Gheorghe COMAN

STATISTICA

Teorie și aplicații

PIM - 2007

Material cu drept de autor

Descrierea CIP a Bibliotecii Naționale a României
COMAN, GHEORGHE
Statistică : teorie și aplicații / Gheorghe Coman. – Iași : PIM, 2007
Bibliogr.
ISBN 978-973-716-729-3

311

Material cu drept de autor

INTRODUCERE

Statistica este ştiința care se ocupă cu analiza cantitativă şi calitativă a fenomenelor de masă.

Statistica presupune observarea fenomenelor de masă la nivelul unui eşantion, în vederea obținerii datelor necesare analizei, prelucrarea acestor date prin intermediul unor metode specifice, formularea unor ipoteze privitoare la fenomenul analizat, testarea acestora şi evidențierea legităților de manifestare a fenomenelor la nivelul ansamblului.

Tehnicile, procedeele şi metodele care servesc demersului statistic sunt reunite în *metodologia* statistică.

În esență, orice ştiință este o metodă, un mod specific de a afla răspunsuri la întrebările pe care ni le punem. Principalele ei caracteristici sunt: **căutarea unor reguli generale (legități), colectarea unor dovezi obiective, operarea cu afirmații controlabile, atitudine sceptică față de cunoştințele acumulate, atitudine deschisă față de orice informații noi, creativitate şi transparență**.

Utilizarea statisticii în ştiință este impusă de exigențele metodei ştiințifice ca metodă de culegere, prelucrare şi interpretare a datelor. Dincolo de procedurile care o compun, metoda statistică este un concept abstract în care ne fundamentăm cunoştințele şi convingerile în viața de fiecare zi. În acest sens, se pot distinge trei modalități de fundamentare:

Tradiționalismul sau argumentul autorității. Ceea ce ştim şi învățăm din contextul social sau cultural se bazează pe obişnuințe sau superstiții. Dacă la baza unei informații cu care operăm se află şi o figură autoritară (părinte, profesor, „cei care au văzut"), atunci atitudinea necritică este şi mai consistentă.

Raționalismul. Baza acestuia este deducția, pornirea de la un principiu general pentru a se ajunge la anumite concluzii specifice. Acest tip de raționament nu aduce un plus de cunoaştere deoarece se bazează pe adevărurile care au condus, în trecut, la constituirea principiului pe care s-a bazat raționamentul deductiv. Spre deosebire de acesta, raționamentul inductiv urmează drumul de la specific la general şi permite noi explorări ale unui subiect. **Inducția este baza metodei ştiințifice în analiza statistică**.

Empirismul. Această modalitate se bazează pe concluzii extrase din experiența de zi cu zi, din observația directă a faptelor.

Modalitățile cunoaşterii comune, enunțate mai sus, nu sunt prin ele însele lipsite de valoare. În acest proces, statistica nu face decât să pună la dispoziție un set de proceduri de calcul şi de raționamente decizionale cu privire la semnificația datelor de cercetare. **Rolul statisticii este acela de a descrie, de a face predicții şi de a conferi credibilitate datelor de observație**. Statistica pune ipoteza cercetării în fața testului negației, prin raportarea la un model aleator de distribuție a valorilor măsurabile.

În general, statistica în ştiințele social-economice se preocupă de trei mari aspecte: **a**) modul în care datele sunt culese, sau mai exact modul

Material cu drept de autor

în care se face selecția observațiilor care urmează sa fie analizate, **b**) descrierea acestor date şi **c**) explicarea şi/sau predicția unor fenomene studiate.

Înainte însă de a trece la tratarea pe larg a acestor aspecte este nevoie de introducerea unor concepte de bază, precum şi a câtorva distincții. Una dintre acestea, şi cea mai importantă poate, este aceea între statistica pur descriptivă şi statistica inferențială. Practic, aceasta din urmă constituie "nucleul dur" al statisticii. Pentru a înțelege distincția mai sus amintită, e nevoie însă mai întâi de clarificarea conceptelor de **populație (statistică)** şi **eşantion**.

Populația statistică reprezintă mulțimea tuturor obiectelor sau indivizilor care prezintă interes pentru studiu. În statistică, când ne referim la populație avem în minte mulțimea unităților de analiză, indiferent ce reprezintă acestea (şcoli, oraşe, întreprinderi, țări, oameni sau chiar procese, fenomene şi acțiuni).

Un eşantion este un subset sau o submulțime a populației analizate. Extragerea unui eşantion din populație este utilă şi chiar necesară în condițiile în care resursele (financiare, de timp etc.) de care dispun inițiatorii studiului nu sunt suficiente pentru a asigura investigarea întregii populații. Un subset din populația analizată, extras conform unei scheme de eşantionare riguroase, poate furniza toată informația necesară la un nivel de acuratețe foarte ridicat.

Putem acum clarifica distincția dintre **statistica descriptivă** şi cea inferențială: ***metodele statistice descriptive constau în descrierea sintetică a informației cuprinse într-un set de date***, iar ***metodele statistice inferențiale constau în acele tehnici şi proceduri folosite pentru a face generalizări despre caracteristicile unei populații, pe baza informațiilor culese de la un eşantion extras din acea populație***. Practic marea provocare a statisticii o constituie exact acest proces de inferență (generalizare) de la datele de eşantion la populație.

Se vor introduce acum două noi concepte, acelea de **parametri** şi respectiv **statistici la nivel de eşantion**.

Caracteristicile populației despre care se face inferențe pe baza eşantionului se numesc parametrii. Caracteristicile eşantionului pe baza cărora inferăm se numesc pur şi simplu statistici. Cele mai multe studii sunt însă interesate în aflarea parametrilor care, în general, sunt necunoscuți şi de alte caracteristici ale populației statistice.

Eşantioanele şi statisticile descriptive sunt utile în măsura în care ele pot oferi informații despre parametrii de interes. Statistica inferențială este aceea care permite obținerea unei măsuri a acurateței statisticilor folosite pentru **estimarea** valorii parametrilor.

Se vor defini acum unele noțiuni deosebit de importante pentru studiul statisticii, şi anume asupra **variabilelor**. **Variabila constituie orice caracteristică a elementelor unei populații sau a unui eşantion care variază (în respectiva populație/eşantion)**.

Material cu drept de autor

Variabilele pot fi clasificate în funcție de multe criterii. Una din distincțiile importante este aceea dintre **variabile discrete** şi **variabile continue**. Atât variabilele discrete cât şi variabilele continue pot lua o infinitate de valori. Diferența dintre ele constă în faptul că în timp ce în cazul variabilelor continue între două valori succesive ale variabilei pot exista o infinitate de valori, în cazul variabilelor discrete acest lucru nu se întâmplă. Un exemplu de variabilă continuă este înălțimea clădirilor unui oraş măsurată în metri, iar un exemplu de variabilă discretă îl reprezintă veniturile indivizilor dintr-o populație, măsurate în lei. În cazul primei variabile, între doua valori succesive ale acesteia (de exemplu 5 şi 6 m) există o infinitate de alte valori deoarece metrii se subdivid în centimetri, apoi în milimetri etc., în cazul veniturilor acest lucru nu mai este posibil, între 5 lei şi 6 lei ne mai existând subdiviziuni.

Măsurarea variabilelor. În esență, **a măsura înseamnă a atribui numere sau simboluri unui aspect al realității obiective sau subiective, în funcție de anumite aspecte cantitative sau calitative care le caracterizează**. În acest mod relația dintre numere sau simboluri ajunge să reflecte relația dintre caracteristicile cărora le-au fost atribuite. **Modul în care sunt atribuite numere sau simboluri pentru a măsura ceva, se numeşte „scală de măsurare"**.

Scopul oricărei măsurări este, într-un fel sau altul, mai direct sau mai puțin direct, acela **de a trage concluzii şi de a susține raționamente**. De aceea, conştientizarea procesului de măsurare este importantă pentru:

- **Cunoaşterea tipurilor de transformări la care putem spune în mod legitim valorile rezultate prin măsurare**. De exemplu, dacă am măsurat distanța în centimetri, ştim că o putem transforma în *inch* prin aplicarea unei reguli, fără a altera semnificația valorilor.
- **Evitarea concluziilor lipsite de sens**. De exemplu, dacă azi sunt afară 20 de grade C şi ieri au fost doar 10, nu putem spune că azi este de două ori mai cald, ci că este cu 10 grade mai cald decât ieri.
- **Alegerea procedurilor statistice adecvate datelor numerice şi scopurilor pe care ni le propunem**.

Niveluri de măsurare a variabilelor

	a. ordonare	b. unitate de măsură	c. zero absolut
Nominal	nu	nu	nu
Ordinal	da	nu	nu
De interval	da	da	nu
De raport	da	da	da

Nivelul de măsurare al variabilelor este un alt criteriu de clasificare a acestora, de o mare importanță pentru studiul statisticii. Putem distinge între patru niveluri de măsurare (**nominal, ordinal, de interval** şi **de**

Material cu drept de autor

raport), în funcție de trei criterii: **a**) posibilitatea de a ordona valorile variabilei; **b**) egalitatea intervalelor dintre valorile variabilei (sau altfel spus existența unei unități de măsură); **c**) existența unei "origini" a variabilei sau, cu alte cuvinte, a unui "zero absolut".

1. Nivelul de măsurare **nominal** presupune clasificarea unor atribute, caracteristici, fenomene etc. în categorii care trebuie să fie distincte, mutual exclusive şi exhaustive. Valorile de tip nominal pot fi, la rândul lor, de două feluri:

- **De identificare**, atunci când o valoare are rolul de codificare a identității, referindu-se în mod unic la o anumită persoană (codul numeric personal, sau un număr de identificare în cadrul unui experiment psihologic etc). Această formă este nerelavantă din punct de vedere propriu-zis statistic, dar este extrem de utilă ca variabilă ajutătoare în manipularea şi organizarea datelor pentru prelucrare.
- **Categoriale**, atunci când desemnează forme pe care le ia o variabilă (tipul de liceu absolvit: „teoretic", „industrial", „artistic"; tipurile temperamentale: „sanguin", „coleric", „flegmatic", „melancolic", etc.). Această formă este în mod obişnuit întrebuințată în psihologie, ori de câte ori este necesară repartizarea subiecților în diverse clase sau categorii, în funcție de prezența sau absența anumitor caracteristici.

Valorile măsurate pe o scală de tip nominal au un caracter calitativ şi nu suportă operații numerice, altele decât cele de sumarizare (numărare, procente).

Acest tip de variabile (respectiv scalele folosite în măsurare) indică numai faptul că exista o diferență calitativă între categoriile studiate, nu şi magnitudinea acestei diferențe. La limită, putem privi aceste variabile ca pe nişte tipologii. Câteva exemple de variabile măsurate la nivel nominal sunt: statutul ocupațional al indivizilor (agricultor, salariat, mic întreprinzător, şomer etc.), religia (ortodox, romano-catolic, greco-catolic etc.), mediul de rezidență (rural, urban) ş.a.m.d. Valorile acestui tip de variabile nu pot fi ordonate, sau cu alte cuvinte nu există o ierarhie (decât eventual conform unor criterii extrinseci) şi în consecință problema "distanței" sau a intervalelor dintre valori nici nu poate fi pusă. Cu atât mai puțin putem discuta despre existența unui "zero absolut" (exemplu: fiecare individ are un statut ocupațional sau aparține unei etnii, sau altfel spus absența caracteristicilor "statut ocupațional" sau "apartenență etnică" este imposibilă).

2. Nivelul de măsurare **ordinal** implică nu numai clasificarea elementelor în categorii ci şi posibilitatea ordonării acestora de la minim la maxim (existența tranzitivității: dacă **a>b** şi **b>c**, atunci **a>c**). Totuşi, la acest nivel de măsurare nu este oferită nici o informație cu privire la "distanța" dintre valorile scalei de măsură. Cu alte cuvinte, diferența dintre prima valoare şi cea de-a doua poate fi diferită de diferența dintre a patra şi a cincia. Exemple de variabile măsurate la nivel ordinal sunt calificativele şcolare (cu valorile "insuficient", "suficient", "bine" şi "foarte bine"), satisfacția față de anumite aspecte (cu valorile "foarte nesatisfăcut", "nesatisfăcut", "satisfăcut", "foarte satisfăcut") etc.

Material cu drept de autor

Rezultă că valorile plasate pe o scală de tip ordinal au o anumită semnificație cantitativă. O anumită valoare este "mai mare" sau "mai bună" decât alta, aflată sub ea. Implicit, ea poate fi "mai mică" sau mai "puțin bună" decât altă valoare, aflată deasupra ei. Dacă o anumită persoană este mai preferată decât alta, şi atribuim primei valoarea 1 iar celei de-a doua valoarea 2, atunci cele două valori se exprimă pe o scală de tip ordinal, care indică doar ordinea preferinței şi nu măsura intensității acestei preferințe. Să ne imaginăm că am avea, pe aceeaşi scală de evaluare, un număr de 6 indivizi. Cel care s-ar plasa pe scala de preferințe pe poziția a 6-a, nu ar fi de şase ori mai preferat ci doar pe a şasea poziție pe scala de preferințe. Un alt exemplu ilustrativ ar putea fi evaluarea satisfacției profesionale pe o scală cu 10 trepte, unde 10 ar fi nivelul de satisfacție cel mai ridicat.

3. Măsurarea la nivel **de interval**, oferă în plus fața de nivel anterior (cel ordinal) şi informație referitoare la distanța dintre valorile scalei şi este caracterizată de existența unor intervale egale. Totuşi, la acest nivel de măsurare nu există un zero absolut, ci mai degrabă unul convențional. O variabilă măsurată pe o scală de interval ne oferă informații nu doar despre ordinea de mărime ci şi despre „dimensiunea" exactă a caracteristicii măsurate. Valorile de acest tip au un caracter cantitativ, exprimat numeric, iar intervalele dintre ele sunt egale.

Exemple:

• temperatura, măsurată pe o scală Celsius. Dacă într-o zi se măsoară 5 grade iar în ziua următoare 10 grade, se poate spune cu precizie că a doua zi a fost cu 5 grade mai cald;

• coeficientul de inteligență măsurat, să zicem, prin numărul de răspunsuri corecte la un test. În acest caz, un rezultat de 30 de răspunsuri corecte este cu 10 unități mai mare decât 20 sau cu 5 unități mai mic decât 35;

Ceea ce este caracteristic valorilor măsurate pe scală de interval este absența unei valori 0 absolute. Cu alte cuvinte, valorile de acest tip nu ne permit evaluări de genul: „O temperatură de 10 grade este de două ori mai mare decât una de 5 grade" sau, „O persoană care a obținut un scor de 30 de puncte este de două ori mai inteligentă decât una care a obținut 15 puncte". Aceasta, deoarece nici temperaturile măsurate pe scala Celsius şi nici inteligența nu au o valoare 0 absolută (dacă acceptăm că nici un om viu nu are inteligență nulă).

4. Măsurarea la nivel **de raport** include toate caracteristicile nivelurilor anterioare (ordonare şi intervale egale), plus existența unei "origini" sau zero absolut. Acest lucru permite formularea unor afirmații în termeni de proporții (raporturi) între valori. La fel ca şi valorile măsurate pe scale de interval, valorile măsurate pe scală de raport suportă toate transformările matematice posibile. Din acest motiv, în practică, valorile măsurate pe scală de interval sau de raport sunt considerate similare, fiind prelucrate prin acelaşi gen de proceduri statistice. Ca urmare, în acest caz, se spune că o variabilă este măsurată pe o „scală de interval/raport".

Material cu drept de autor

De exemplu, vitezele de răspuns a doi subiecți la un acelaşi stimul pot fi comparate în termeni de "timpul de răspuns a fost de două ori mai mare" etc.. Exemple de variabile măsurate la acest nivel sunt vârsta, greutatea, înălțimea, distanța, numărul de copii din gospodărie etc.

Dacă luăm în considerare proprietățile numerice şi tipul de transformări suportate de fiecare scală de măsurare, atunci ordinea crescătoare a acestora este **nominal – ordinal – interval - raport**. Din acest punct de vedere se poate chiar spune că scalele de măsurare se plasează pe o scală ordinală.

Corecta identificare a nivelului de măsurare utilizat este foarte importantă în alegerea procedurilor statistice de analiză. După cum se poate observa din descrierea de mai sus, pentru fiecare nivel exista operații matematice permise şi operații interzise. Astfel, la primul nivel, cel nominal nu sunt permise nici ordonarea, nici adunarea/scăderea şi nici înmulțirea/împărțirea. La nivelul ordinal este permisă numai ordonarea, la cel de interval sunt permise în plus şi operațiile de adunare/scădere, iar la ultimul nivel, cel de raport sunt permise toate operațiile.

În funcție de nivelul de măsurare, se poate vorbi despre variabile măsurate la nivel nominal, variabile măsurate la nivel ordinal etc., sau, mai pe scurt, variabile nominale, ordinale, de interval şi de raport. Reducând cele patru clase la două, se poate vorbi de *variabile calitative* (nivelurile nominal şi ordinal) şi *variabile cantitative* (interval şi raport). Datorita caracterului "ierarhic" şi cumulativ al nivelurilor de măsurare (de la multe restricții către nici o restricție în ceea ce priveşte operațiile permise, sau de la "calitativ" la "cantitativ"), vom putea întotdeauna trata o variabilă aflată la un nivel "superior" de măsurare ca şi cum ar fi fost măsurată la un nivel "inferior". De exemplu, vârsta măsurata în ani de viață va putea oricând fi tratată ca o variabilă ordinală, dacă îi grupăm valorile (sub 20, 21-30, 31-50, peste 50). Niciodată însă nu vom putea trata o variabilă aflată la un nivel "inferior" ca pe una aflată "mai sus" în ierarhie.

Material cu drept de autor

Cap.1. VARIABILE STATISTICE

1.1. Concepte de bază folosite în statistică

Principalele concepte folosite în statistică sunt următoarele: 1. colectivitatea (populația) statistică; 2. unitatea statistică; 3. variabila (caracteristica) statistică; 4. indicatorul statistic.

1. **Colectivitatea statistică**, numită şi populație statistică, reprezintă masa totală sau globală a evenimentelor distincte din cadrul unui fenomen sau proces natural sau socio-economic ,supus cercetării statistice.

Colectivitatea este de doua feluri: colectivitate totală; colectivitate parțială.

Colectivitatea totală reprezintă situația în care toate aparițiile fenomenelor şi proceselor cu aceeaşi caracteristică studiată sunt supuse cercetării .

Colectivitatea parțială reprezintă situația în care doar o parte din manifestările individuale vor intra sub incidența cercetării statistice. Ambele colectivități pot fi *statice* sau *dinamice*, în funcție de variabilitatea lor în timp.

2. Prin **unitate statistică** se înțelege **entitatea componentă a unei colectivități, purtătoare a unei însuşiri care o face interesantă studiului statistic**.

3. **Variabilă statistică**, numită şi **caracteristică statistică**, este **caracteristica care poate prezenta variație valorică de la o unitate statistică la altă unitate statistică al unei colectivități statistice** fiind *atributul sau însuşirea pe care o are unitatea statistică supusă cercetării statistice*. Este de mai multe tipuri: *variabilă de timp* (an, lună, zi, etc.); *variabilă de spațiu* (unitățile administrativ-teritoriale); *variabilă atributivă*. Formele concrete de manifestare a caracteristicilor la nivelul unei unități statistice se numesc ***variante*** sau valori. Variabila statistică se reprezintă printr-un simbol care poate lua orice valoare în domeniul de definiție al acesteia.

Variabilele alternative, atributive sau **dihotomice** sunt cele care au doar două posibilități de reprezentare (da-nu, 0-1, alb-negru), ele grupându-se la rândul lor în variabile: *numerice; nenumerice (alfanumerice)*.

Variabilele nealternative acoperă întreaga gamă de trăsături şi caracteristici calitative ale fenomenelor care pot fi studiate cantitativ. Şi ele la rândul lor se clasifică în variabile *numerice* şi *nenumerice*.

Variabilă aleatoare este caracteristica a cărei valoare este supusă întâmplării, incertitudini sau imprevizibilului în domeniul de definiție a ei.

4. **Indicatorii statistici** sunt o măsură numerică a variabilelor statistice şi reprezintă datele statistice cu ajutorul cărora se cercetează un fenomen sau proces economic sau social sub raportul structurii, interdependențelor, al modificării lor în timp şi spațiu.

Indicatorii statistici sunt de două tipuri: **indicatori primari**; **indicatori derivați**.

Material cu drept de autor

Indicatorii primari sunt extraşi direct din realitatea sub toate formele ei de organizare.

Indicatorii derivaţi sunt obţinuţi prin transformarea logico-matematică a indicatorilor primari , ei găsindu-se sub următoarele forme: *indicatori absoluţi; indicatori relativi; mediile statistice; indicii statistici; ecuaţiile de estimare.*

Proces statistic. Este setul de condiţii, stări sau operaţii care în mod repetabil vin împreună să transforme intrările în ieşiri.

Informaţia statistică. Reprezintă conţinutul specific (semnificaţia), mesajul datelor. Datele statistice se întâlnesc adesea ca indicatori statistici sau parametri statistici.

Parametri statistici reprezintă expresia numerică care însumează, sintetizează câteva aspecte ale populaţiei statistice sau caracteristicile unui proces. Se mai numeşte şi valoare tipică. În funcţie de conţinutul lor se disting parametrii de nivel (media, mediana, modulul), parametrii de variaţie (dispersia, abaterea standard), coeficienţi de asimetrie, ş.a.

Inferenţa statistică este procedura de analiză şi cunoaştere a informaţiei de grup (parte, eşantion) folosită pentru a cunoaşte populaţia întreagă.

Confidenţă este intervalul în care rezultatul unei analize statistice realizate pe o parte a populaţiei statistice este corect pentru populaţia întreagă sau eronat dar nu cu mai mult decât un nivel dat.

Model statistic este construcţia logică (funcţia, ecuaţia, sistemul) care exprimă trăsăturile şi corelaţiile esenţiale din manifestarea reală a fenomenului sau procesului studiat.

1.2. Variabile dependente şi variabile independente

În esenţă, un studiu statistic îşi propune evidenţierea legăturilor dintre diverse caracteristici ale realităţii (variabile). În acest context există variabile ale căror valori sunt **dependente** pentru că variază în funcţie de valorile altei sau altor variabile, care sunt denumite, din acest motiv, **independente**. Identificarea lor corectă în cazul unui studiu statistic este esenţială pentru fundamentarea procedurilor statistice.

În mod esenţial, variabila dependentă face obiectul măsurării cu scopul de a fi supusă unor concluzii. Prin opoziţie, variabila independentă este utilizată ca variabilă de influenţă, ale căror efecte posibile asupra variabilei dependente urmează sa fie puse în evidenţă. Termenii „dependent", „independent" se utilizează în mod obişnuit în legătură cu cercetarea experimentală. În acest context există variabile „manipulate" adică „independente" de reacţiile, intenţiile, conduitele sau trăirile subiecţilor investigaţi (toate acestea fiind variabile „dependente"). În raport cu analiza statistică, definirea variabilelor ca dependente şi independente nu este condiţionată de măsurarea lor în condiţii de experiment.

Vom reţine faptul că nu există variabile care sunt „dependente" sau „independente" prin natura lor. Caracteristica de a fi de un tip sau de altul

Material cu drept de autor

provine din rolul care le este atribuit de către cercetător într-un anumit context de cercetare.

Între variabila independentă şi variabila dependentă se formează ceea ce se numeşte o **relație statistică** şi care reprezintă expresia matematică ce arată cum o variabilă este relaționată cu una sau mai multe variabile ignorând pentru un timp efectele factorilor minori în sistem.

Dacă pentru fiecare valoare pe care o poate lua o variabilă **x** îi corespunde una sau mai multe valori ale unei alte variabile **y** spunem că **y** este o funcție a lui **x** adică: **y = f(x)**. Variabila **x** se numeşte variabilă independentă (factorială, cauză) iar variabila **y** se numeşte dependentă (rezultativă, efect).

Dacă pentru fiecare valoare a lui **x** corespunde o singură valoare a lui **y** atunci spunem că **y** este o funcție univocă a lui **x**. Altfel funcția este neunivocă. De asemenea funcția poate să fie liniară sau neliniară. Şi pentru că în economie o variabilă depinde de mult mai multe alte variabile decât cele luate în calcul forma funcției statistice va fi **y = f(x)** + ε; ε exprimă influența factorilor minori, nesemnificativi, neluați în analiză.

Material cu drept de autor

Cap.2. PROBABILITATE ŞI DISTRIBUȚII DE PROBABILITATE

2.1. Noțiunea de probabilitate

În sens curent şi epistemologic, cuvântul „**probabilitate**", provenit din latinescul „probabilitas", înseamnă **credibilitate, verosimilitate**. În sens matematic, probabilitatea este un calcul absolut sigur prin el însuşi. Dar, epistemologic, probabilitatea se opune certitudinii[1]. Referindu-se la noua logică, probabilistă, Immanuel Kant scria: ***"Teoria certitudinii cunoştinţelor noastre cuprinde şi teoria despre cunoaşterea probabilului, care trebuie considerat o aproximare a certitudinii. Prin probabilitate trebuie să înţelegem un asentiment bazat pe motive insuficiente, care este însă legat mai mult de motivele suficiente decât de cele ale contrariului"***. Însă ***"...numai matematicianul poate să determine legătura dintre motivele suficiente şi cele insuficiente"*** întrucât ***"probabilitatea este o aproximare a certitudinii. În cazul probabilităţii trebuie să existe întotdeauna un etalon prin care s-o pot aprecia. Acest etalon este certitudinea"***[2].

Teoria probabilității poate să analizeze acele situații în care nu avem suficiente informații care să permită aplicarea logicii clasice; ea este capabilă să ne dea cel mai bun tip de răspuns pe care-l justifică o informație incompletă

Raționamentul probabilistic completează modul de a gândi al logicii clasice cu raționamentul inductiv.

Gândirea inductivă este lipsită de rigoare. Dar există în ştiința modernă anumite metode de analiză a concluziilor inductive acordând un anumit grad de credibilitate acestora, stabilite de calculul probabilităților şi statistica matematică. Pentru a stabili regulile de precizare a gradului de credibilitate a concluziilor inductive vom sublinia deosebire profundă dintre raționamentul deductiv şi raționamentul inductiv. Cel mai bun exemplu în acest sens îl prezintă stabilirea legilor termodinamicii. Ele n-au fost demonstrate nici odată, însă sunt unanim acceptate de ştiință întrucât nici nu au fost infirmate vreodată.

De la raționamentul inductiv s-a ajuns la concluziile statistice. Următoarele principii caracterizează unele proprietăți esențiale ale raționamentului inductiv care au înlesnit trecerea la raționamentul probabilistic şi, deci, la statistică.

1. *Concluziile* bazate pe argumentația inductivă sunt posibile numai cu oarecare *probabilitate* şi niciodată cu siguranță deplină. Din însăşi definiția inducției rezultă că aceste concluzii sunt aplicate pentru cazurile

---

[1]Elisabeth Clement ş.a., *Filosofia de la A la Z*, Bucureşti, Editura ALL Educational, 2000, p.417-418.

[2]Immanuel Kant, *Logica generală*, Editura Ştiințifică şi Enciclopedică, Bucureşti, 1985, p.136-137

Material cu drept de autor

care nu au fost cuprinse în numărul celor observate. Când toate cazurile asupra cărora este extinsă concluzia sunt incluse în sfera observațiilor, concluzia încetează să fie un rezultat al inducției. În mod corespunzător, deşi inducția constituie în mare, o măsură, o metodă eficace pentru lărgirea cunoştințelor omului, ea este legată totdeauna de un risc. Aplicarea concluziilor la cazurile care nu au intrat însă în sfera observării se poate asemăna totdeauna cu un salt necunoscut.

2. Raționamentul inductiv impune necesitatea de a ne referi la elemente care se găsesc în afara faptelor din premise. Noi nu avem de-a face cu un sistem limitat ci cu unul extins din care numai o parte poate fi cercetată direct, iar în concluzie intră şi o mare parte de cazuri care n-au fost supuse observării. Nu totdeauna faptele cuprinse în premise au relații apropiate cu concluzia noastră. Atât timp cât nu există posibilitatea să se ia în considerație toate elementele care se referă la concluzia inductivă, acei care formulează concluzii de ordin inductiv exprimă o capacitate de pătrundere şi perspicacitate în stabilirea unor previziuni corecte privind fenomenele analizate.

3. Trebuie să presupunem că există o oarecare uniformitate în sistemul faptelor din care fac parte premisele şi deducțiile concluziilor inductive; aceasta constituie justificarea logică a saltului în necunoscut care se găseşte totdeauna în inducție. Dacă nu ar fi existat o asemenea uniformitate în procesele naturale, dacă natura ar fi un adevărat haos, nici o îngrămădire de fapte nu ar fi putut justifica inducția.

4. Verificarea inducției necesită confirmarea obiectivă. Legitatea formală a inducției (de exemplu: lanțul deducțiilor matematice) se bazează în întregime pe concordanța interioară. Adevărul matematic reprezintă lipsa de contradicții. Deducțiile concluziilor inductive pot fi verificate definitiv numai prin observare şi dacă ele sunt juste nu trebuie să contrazică faptele reale în domeniul dat.

Acest mod de gândire ajunge la un nivel rafinat în matematică. Precizia sa depinde aici de mai mulți factori: modul precis în care sunt definiți termenii, rigoarea cu care definițiile sunt respectate şi atenția cu care toate regulile de acțiune sunt puse în evidență şi clar exprimate.

O concluzie logică poate să merite pe drept adjectivele corectă, sănătoasă sau precisă, toate acestea însemnând că ea a fost dedusă în mod convenabil din materialul inițial. Dar faptul că a fost obținută prin metode logice corecte nu înseamnă câtuşi de puțin că ea este în mod necesar adevărată. Dacă s-a ştiut într-un fel sau altul că afirmațiile inițiale sunt adevărate, atunci consecințele logice deduse trebuie să fie şi ele adevărate. Dacă însă suntem interesați să aflăm adevărul, atunci sau trebuie să o luăm de la început şi să stabilim adevărul afirmațiilor inițiale sau, neglijând procesul logic prin care le-am obținut, să stabilim adevărul propozițiilor dobândite prin inferență, folosind o metodă direct aplicabilă lor.

Domeniul de studiu al teoriei probabilităților şi statisticii matematice îl formează legile ce se manifestă în domeniul fenomenelor întâmplătoare cu caracter de masă, din natură şi viața economico-socială. Aceste legi poartă

Material cu drept de autor

denumirea de ***legități probabilistice (stochastice)*** sau ***legi de probabilitate***; ele au un caracter obiectiv, necesar şi general.

Matematic, probabilitatea este definită prin relațiile: funcția **P(E) → [0, 1]** cu proprietățile: **p(E) = 1**; şi

$$p(A \cup B) = p(A) + p(B), (\forall) A, B \in p(E) si\ A \cap B = \varnothing .$$

Probabilitatea se apreciază pe baza frecvenței relative a evenimentelor aleatoare. ***Prin frecvență relativă a evenimentului A se înțelege raportul dintre numărul probelor n_A în care evenimentul A s-a produs şi numărul total n de probe (experiențe) efectuate***. Notând prin **$f_n(A)$** frecvența relativă, se va obține:

$$f_n(A) = \frac{n_A}{n} \tag{2.1}$$

Numărul **n_A** care exprimă de câte ori s-a produs evenimentul **A**, în cele **n** experiențe, se numeşte4 ***frecvență absolută*** a evenimentului **A**. Din expresia (2.1) rezultă:

$$0 \le n_A \le n \ \text{şi} \ \ 0 \le f_n(A) \le 1 \tag{2.2}$$

Dintr-o îndelungată observație a fenomenelor şi proceselor de masă s-a putut constata că dacă un experiment aleatoriu se repetă, în aceleaşi condiții, de un număr suficient de mare de ori, atunci frecvența relativă primeşte o anumită stabilitate, oscilând în jurul unui număr **p – probabilitatea evenimentului A**. Tocmai de aceea, drept măsură cantitativă de apreciere a posibilității obiective de a se produce evenimentul întâmplător **A** poate fi luată frecvența relativă **$f_n(A)$**, rezultată după un număr **n** suficient de mare de experiențe, efectuate în aceleaşi condiții.

Definiția clasică a probabilității. Să presupunem că spațiul de selecție asociat unei experiențe aleatoare este finit, adică **E = {e_1, e_2, ...,e_n}** şi, mai mult, cele **n** evenimente elementare $e_i, i = \overline{1,n}$, sunt egal probabile (egal posibile, adică au aceeaşi şansă de realizare).

Dacă **A** este un eveniment ce poate să apară ca rezultat al acestui experiment aleatoriu, atunci, evident, **A** aparține unui câmp de evenimente **K** generat de experimentul aleatoriu considerat. Mai mult, un asemenea eveniment aleatoriu **A** poate fi exprimat, în mod unic, ca şi o reuniune de evenimente elementare ale spațiului de selecție, deci sub forma:

$$A = e_1 \cup e_2 \cup ... \cup e_k \tag{2.3}$$

unde $e_i, i = \overline{1,k}$, sunt cele **k** evenimente elementare, egal probabile, care favorizează apariția evenimentului **A**.

Se poate da următoarea definiție. ***Se numeşte probabilitate a unui eveniment A şi se notează P(A), raportul dintre numărul evenimentelor elementare favorabile evenimentului A şi numărul total de evenimente n elementare egal posibile, adică***:

Material cu drept de autor

$$P(A) = \frac{k}{n} \qquad (2.4)$$

Cu alte cuvinte, dacă într-o operație de masă, care are loc în condiții identice, un eveniment **A** se produce în medie de **k** ori, adică la **k** din **n** elemente ale unei colectivități, atunci relația (2.4) sintetizează definiția noțiunii de probabilitate: ***se numeşte probabilitate a unui eveniment, raportul dintre numărul k de rezultate favorabile producerii evenimentului A şi numărul total n de rezultate posibile ale experimentului aleatoriu, cu condiția ca toate rezultatele să fie egal posibile***.

De remarcat este faptul că probabilitatea unui eveniment are sens atât timp cât condițiile în care are loc operația de masă respectivă rămân neschimbate. Orice schimbare a acestor condiții atrage după sine şi schimbarea probabilității.

Probabilitatea **P** prezintă următoarele proprietăți evidente:

1^o. P(∅) = 0;

2^o. P(E) = 1;

3^o. P(A) ∈ **[0, 1]** pentru orice **A** ∈ **K.**

O interpretare firească a acestor proprietăți revine la: cu cât **P(A)** este mai apropiată de **1**, cu atât evenimentul **A** are loc mai des; dacă **P(A) = 1**, evenimentul **A** are loc totdeauna, deci este un eveniment sigur. Dimpotrivă, dacă **P(A)** este foarte mică, atunci evenimentul **A** are loc foarte rar, iar dacă **P(A) = 0**, atunci evenimentul **A** se poate considera ca un eveniment imposibil. Dacă **A** este un eveniment aleatoriu (întâmplător), atunci, evident, **P(A)** este cuprins între **0** şi **1**.

Definiția clasică a probabilității se bazează pe două ipoteze:

- într-o operație de masă, deci rezultatul unui experiment aleatoriu, este constituit dintr-un număr bine determinat **n** de cazuri egal probabile, care se exclud reciproc;

- evenimentul urmărit **A** se produce în **k** cazuri (numite favorabile lui **A**) şi nu se produce în celelalte **n – k** cazuri (numite nefavorabile lui **A**).

Această formulare a definiției probabilității unui eveniment a fost dată de Laplace în lucrarea sa fundamentală *Teoria analitică a probabilității*. O definiție cu caracter calitativ - filozofic şi care în esență exprimă şi definiția clasică a probabilității, a fost dată, pentru prima dată, de către Jaques Bernoulli. După Bernoulli, ***probabilitatea este gradul certitudinii, care se raportează la certitudine ca partea la întreg***, iar Immanuel Kant scria: ***"probabilitatea este o aproximare a certitudinii. În cazul probabilității trebuie să existe întotdeauna un etalon prin care s-o pot aprecia. Acest etalon este certitudinea"***.

Definiția empirică a probabilității. Critica definițiilor anterioare a condus ca la începutul secolului XX, Richard von Mises, care este considerat creatorul teoriei frecvențelor, a definit probabilitatea ca limită a frecvenței relative, atunci când fenomenul ar fi supus la un număr nelimitat de probe, adică:

Material cu drept de autor

$$P(A) = \lim_{n\to\infty} f_n(A) = \lim_{n\to\infty} \frac{n_A}{n} \tag{2.5}$$

în care $\mathbf{f_n}$ este frecvența corespunzătoare pentru **n** probe.

Deşi şi această relație a lui Richard von Mises a fost criticată, ea continuă să fie considerată cea mai bună definiție a probabilității.

2.2. Funcție de repartiție

A descrie o variabilă aleatoare din punct de vedere probabilistic înseamnă a indica exact probabilitatea fiecărei valori posibile adică a fiecărui eveniment. Aceasta înseamnă a-i stabili legea de repartiție (distribuție).

*Se numeşte **funcție** sau **lege de repartiție (distribuție)** a unei variabile aleatoare orice relație care stabileşte o corespondență între valorile posibile ale acestei variabile şi probabilitățile lor.*

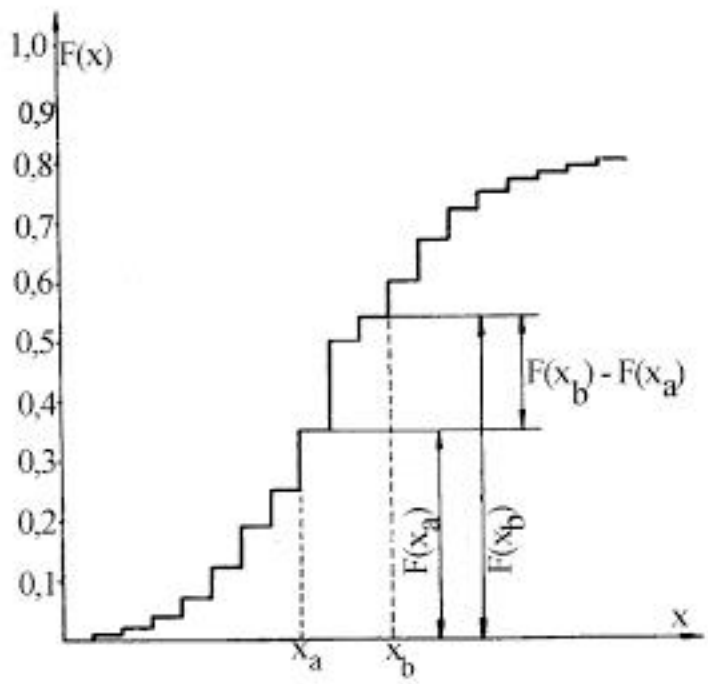
Prin convenție se înseamnă variabila aleatoare cu majusculă şi valorile ei posibile cu literele mici corespunzătoare. De exemplu **X** şi $\mathbf{x_1, x_2, x_3,..,x_n}$.

Legea de repartiție a unei variabile **aleatoare discrete** este dată, în mod obişnuit, printr-un **tabel** care constituie şi notația variabilei:

$$X:\begin{pmatrix} x_1 & x_2 & x_3 & \cdots\cdots & x_n \\ p_1 & p_2 & p_3 & \cdots\cdots & p_n \end{pmatrix};\ p_i = P(X = x_i);\ \sum_{i=-\infty}^{+\infty} p(x_i) = 1 \tag{2.6}$$

În aplicații practice ne interesează adesea probabilitatea ca valorile variabilei aleatoare să fie mai mici decât o valoarea dată $\mathbf{x_k}$. Această condiție apare sub forma:

$$P(X \le x_k) = \sum_{i=-\infty}^{k} p(x_k) \tag{2.7}$$



După cum se observă din (2.7), probabilitatea $P(X \le x_k)$ depinde de valorile variabilei întâmplătoare **X**, prin urmare este o funcție de **x**. De aceea, probabilitatea ca valorile variabilei întâmplătoare să fie mai mici sau cel mult egale cu o valoare dată $\mathbf{X = x_k}$ se numeşte ***funcție de repartiție*** şi se notează **F(x)**, adică:

Fig.2.1. Reprezentarea grafică a funcției de repartiție pe variabila aleatoare discontinuă

$$F(x) = P(X \le x_k) = \sum_{i=-\infty}^{k} p(x_i) \tag{2.8}$$

Material cu drept de autor

În figura 2.1 sunt reprezentate valorile funcției de repartiție corespunzătoare valorilor **X = x_i (i = ... -2, -1, 0, 1, 2, 3, ...)**.

Dacă **x_a** şi **x_b** sunt două valori oarecare ale variabilei şi dacă **$x_a < x_b$**, atunci, după cum se observă în figura 2.1:

$$F(x_a) < F(x_b) \tag{2.9}$$

prin urmare funcția de repartiție este nedescrescătoare. Fiind suma unor probabilități, funcția de repartiție este o cantitate nenegativă, adică:

$$0 \le F(x) \le 1 \tag{2.10}$$

adică funcția de repartiție, ca orice probabilitate, este cuprinsă între zero şi unu.

În practică, se pune adesea problema determinării probabilității ca valorile variabilei întâmplătoare **X** să se găsească între două valori date **X = x_a** şi **X = x_b**, unde **a < b**, adică:

$$P(x_a \le X \le x_b) \tag{2.11}$$

Evenimentul **X ≤ x_b** se descompune în evenimentele **X ≤ x_a** şi **x_a < X ≤ x_b**, aşa încât:

$$P(x \le x_b) = P(x \le x_a) + P(x_a < x \le x_b) \tag{2.12}$$

de unde:

$$P(x_a < x \le x_b) = P(x \le x_b) - P(x \le x_a) \tag{2.13}$$

care în baza relației (2.8) se poate scrie:

$$P(x_a < X \le x_b) = F(x_b) - F(x_a) \tag{2.14}$$

dar:

$$F(x_b) = \sum_{i=-\infty}^{b} p(x_i), \qquad \text{iar} \qquad F(x_a) = \sum_{i=-\infty}^{a} p(x_i)$$

aşa încât:

$$P(x_a < X \le x_b) = \sum_{i=-\infty}^{b} p(x_i) - \sum_{i=-\infty}^{a} p(x_i) \tag{2.15}$$

Funcția de repartiție este caracteristica cea mai universală a unei variabile aleatoare, fie discretă fie continuă. Funcția de repartiție caracterizează complet o variabilă aleatoare din punct de vedere probabilistic, adică este una din formele legii de repartiție.

$$F(x) = P(X < x) \tag{2.16}$$

Dar, plecându-se de la funcția de repartiție a unei variabile discrete oarecare care este totdeauna o funcție discontinuă, în scară, cu salturi care se produc în punctele în care sunt valori posibile ale acestei variabile aleatoare şi sunt egale cu probabilitățile acestor valori, figura 2.2, pe măsură ce numărul valorilor posibile ale variabilei aleatoare creşte şi intervalele dintre ele se micşorează, numărul treptelor devine tot mai mare şi treptele

Material cu drept de autor

tot mai mici, după care variabila aleatoare tinde către o variabilă continuă şi funcția sa de repartiție către o funcție continuă, figura 2.3.

Dacă funcția de repartiție (2.16) este continuă şi are o derivată continuă, atunci se spune că atât variabila întâmplătoare **X** cât şi repartiția ei sunt continue.

Derivata funcției de repartiție:

$$f(x) = \frac{dF(x)}{dx} \tag{2.17}$$

se numeşte ***funcție de distribuție*** sau ***funcție de frecvență a distribuției variabilei aleatoare***.

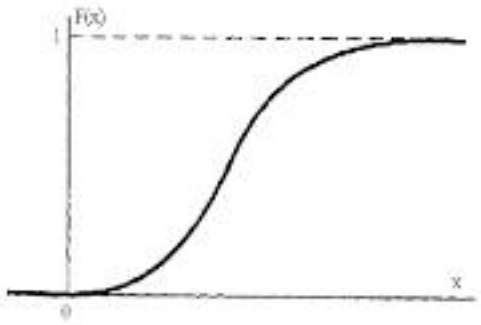


Fig.2.2. Funcția de repartiție cu un număr mare de trepte

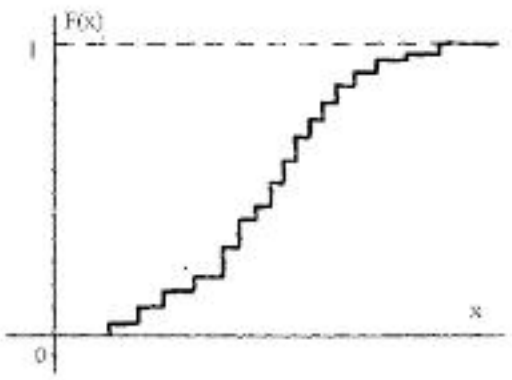


Fig.2.3. Funcția de repartiție continuă

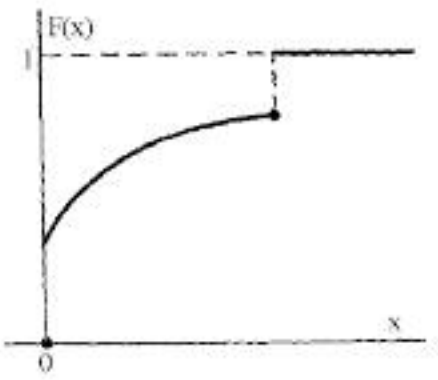


Fig.2.4. Funcție de repartiție continuă, cu o discontinuitate

În general, în aplicațiile practice, funcția de repartiție a unei variabile aleatoare continue este o funcție continuă în toate punctele. Se pot găsi totuşi exemple de variabile aleatoare ale căror valori posibile umplu în mod continuu un anumit interval, dar funcția de repartiție nu este peste tot continuă, admițând discontinuități, figura 2.3. Aceste variabile aleatoare se numesc ***mixte.*** Un exemplu de variabilă aleatoare mixtă este timpul **T** de funcționare fără defect a unui aparat supus încercărilor pe o durată **t**. Funcția de repartiție a acestei variabile aleatoare este continuă peste tot, cu excepția punctului **t**.

În aplicațiile practice este uneori necesar să se calculeze probabilitatea de încadrare unei variabile aleatoare de a lua o valoare între anumite limite, de exemplu între α şi β. Vom numi acest eveniment "***apartenența variabilei aleatoare intervalului cuprins între*** α ***şi*** β**"**.

Pentru a fixa ideile, să convenim că extremitatea stângă a intervalului îi aparține, iar extremitatea din dreaptă nu. Atunci, a cere ca variabila **X** să cadă în intervalul în cauză înseamnă că trebuie să fie verificată inegalitatea:

$$\alpha \leq \mathbf{X} < \beta$$

Material cu drept de autor

Să exprimăm probabilitatea acestui eveniment cu ajutorul funcției de repartiție a lui **X**. În acest scop să considerăm următoarele trei evenimente:

evenimentul **A** , constând în aceea că $X < \beta$;
evenimentul **B**, constând în aceea că $X < \alpha$;
evenimentul **C**, constând în aceea că $\alpha \le X < \beta$.

Cum **A = B + C**, în virtutea teoremei de adunare a probabilităților, avem :

$$P(X < \beta) = P(X < \alpha) + P(\alpha \le X < \beta)$$

sau

$$F(\beta) = F(\alpha) + P(\alpha \le X < \beta)$$

de unde

$$P(\alpha \le X < \beta) = F(\beta) - F(\alpha) \qquad (2.18)$$

adică : **probabilitatea ca variabila aleatoare să cadă într-un anumit interval este egală cu creşterea funcției de repartiție pe acest interval.**

Dacă se micşorează nedefinit domeniul (α, β) presupunând că $\beta \to \alpha$, în loc de probabilitatea pe un anumit interval se obține, la limită, probabilitatea ca variabila să ia o valoare particulară :

$$P(X = \alpha) = \lim_{\beta \to \alpha} P(\alpha \le X < \beta) = \lim_{\beta \to \alpha} [F(\beta) - F(\alpha)] \qquad (2.19)$$

Valoarea acestei limite depinde de continuitatea sau discontinuitatea funcției **F(x)** în punctul $x = \alpha$. Dacă în punctul α funcția **F(x)** are o discontinuitate, limita din (2.19) este egală cu valoarea saltului funcției **F(x)** în acest punct. Dacă, din contra, funcția **F(x)** este continuă în punctul α, această limită este egală cu zero.

Mai departe vom spune că variabilele aleatoare sunt continue numai dacă funcția lor de repartiție este peste tot continuă.

Se poate deci formula proprietatea următoare: ***probabilitatea unei valori oarecare a unei variabile continui este egală cu zero.***

Daca probabilitatea evenimentului $X = \alpha$ este nulă, aceasta nu înseamnă că respectivul eveniment nu va apare ci că frecvența lui este nulă. S-a văzut deja că pentru un mare număr de experiențe frecvența unui eveniment nu este egală ci tinde spre probabilitatea sa. Faptul că probabilitatea evenimentului $X = \alpha$ este nulă înseamnă numai că pentru un număr de experiențe infinit de mare acest eveniment se va produce cât de rar vrem.

Dacă un eveniment **A** are o probabilitate de realizare nulă într-o anumită experiență, evenimentul contrar $\bar{A}$ are o probabilitate egală cu unitatea, totuşi el nu este sigur. Pentru o variabilă aleatoare continuă **X** pentru un α oarecare probabilitatea evenimentului $X \neq \alpha$ este egală cu unitatea, totuşi acest eveniment nu este sigur. Dacă numărul de experiențe se măreşte la infinit, acest eveniment va avea loc aproape totdeauna dar nu totdeauna.

Material cu drept de autor

2.3. Densitatea de probabilitate.

Fie o variabilă aleatoare continuă **X** dată prin funcția sa de repartiție **F(x)** pe care o vom presupune continuă şi derivabilă. Să calculăm, pentru această variabilă, probabilitatea de a se găsi în intervalul **x - (x+Δx)**:

$$P(x<X<x+\Delta x) = F(x+\Delta x) - F(x) \qquad (2.20)$$

deci creşterea funcției de repartiție pe acest interval.

Să considerăm raportul dintre această probabilitate şi lungimea intervalului, adică probabilitatea medie pe unitatea de lungime a intervalului şi să facem ca **Δx** să tindă către zero. La limită se va obține derivata funcției de repartiție:

$$\lim_{\Delta x \to 0} \frac{F(x+\Delta x) - F(x)}{\Delta x} = F'(x) \qquad (2.21)$$

Să introducem notația:

$$F'(x) = f(x)$$

Funcția **f(x)** - ***derivata funcției de repartiție*** - caracterizează densitatea de repartiție a valorilor variabilei aleatoare într-un punct dat. Această funcție se numeşte ***densitate de repartiție a probabilităților*** sau, pe scurt, ***densitate de probabilitate*** a variabilei aleatoare continue **X**. Aceşti termeni devin mai clari dacă se utilizează interpretarea mecanică a unei repartiții; Funcția **f(x)** caracterizează formal densitatea de repartiție a maselor pe axa absciselor ("densitate lineară").

Densitatea de probabilitate, ca şi funcția de repartiție, este una din formele legii de repartiție. Contrar funcției de repartiție, această formă nu este universală, căci ea nu are sens decât pentru variabile aleatoare continui.

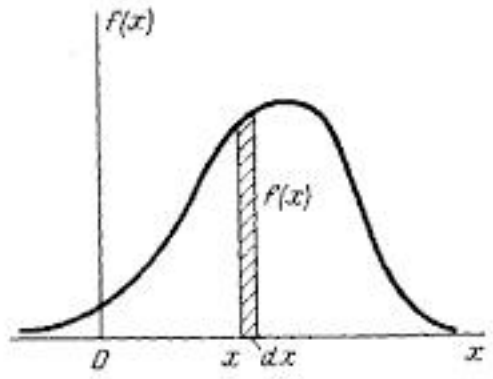


Fig.2.5. Densitatea elementară de probabilitate

Să considerăm o variabilă aleatoare continuă **X** cu o densitate de probabilitate **f(x)** şi domeniul elementar **dx** adiacent punctului **x**, figura 2.5. Probabilitatea pentru variabila aleatoare **X** de a se afla în acest interval elementar este egală cu **f(x).dx**. *Mărimea* **f(x).dx** *se numeşte element de probabilitate*. Din punct de vedere geometric este aria dreptunghiului elementar având ca bază **dx**.

Să exprimăm probabilitatea ca variabila **X** să se afle în intervalul (α,β), figura 2.6, în funcție de densitatea de probabilitate. Este evident că probabilitatea căutată este egală cu suma elementelor de probabilitate pe acest interval, adică integrala următoare:

$$P(\alpha < X < \beta) = \int_{\alpha}^{\beta} f(x)dx \qquad (2.22)$$

Material cu drept de autor

Din punct de vedere geometric probabilitatea pentru variabila aleatoare **X** de a se afla în intervalul (α,β) este egală cu aria cuprinsă între curba densității şi axa absciselor, limitată de ordonatele extremităților intervalului (α,β), figura 2.6.

Formula (2.22) exprimă o densitate de probabilitate cu ajutorul unei funcții de repartiție. Fie acum problema inversă, adică să găsim o funcție de repartiție în funcție de o densitate. Prin definiție:

$$F(x) = P(X < x) = P(-\infty < X < x) \qquad (2.23)$$

şi în virtutea formulei (2.23):

$$F(x) = \int_{-\infty}^{x} f(x)dx \qquad (2.24)$$

Din punct de vedere geometric **F(x)** nu este nimic altceva decât aria cuprinsă între curba densităților şi axa absciselor la stânga punctului **x**, figura 2.7.

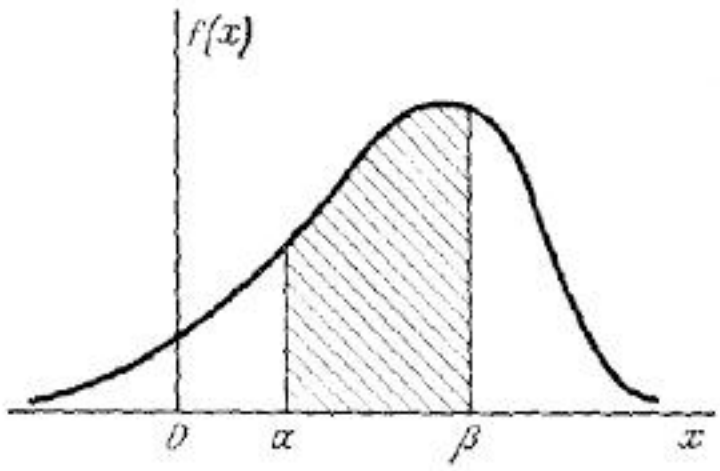


Fig.2.6. Domeniul de probabilitate între α şi β

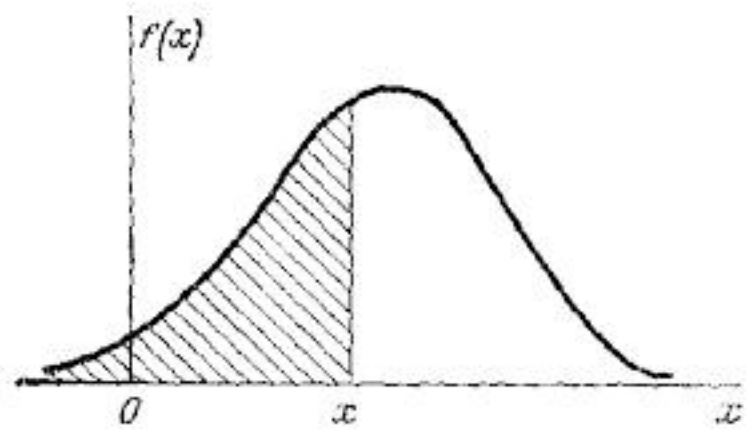


Fig.2.7. Domeniul de probabilitate între **-∞, x**

Să enunțăm proprietățile principale ala densității de probabilitate :

1. Densitatea de probabilitate este o funcție nenegativă:

$$f(x) \geq 0$$

Această proprietate decurge imediat din faptul că funcția de repartiție **F(x)** *este nedescrescătoare.*

2. Integrala intre **-∞** şi **+∞** a densității de probabilitate este egală cu unitatea :

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

Aceasta decurge din formula (2.24) şi din faptul ca **F(+∞) = 1**.

Material cu drept de autor

Ca proprietăți esențiale ale densității de probabilitate se pot da următoarele interpretări geometrice :

1. orice curbă a densităților de probabilitate se găseşte deasupra axei absciselor;
2. aria totală dintre curba de densitate şi axa absciselor este egală cu unitatea

Să studiem acum dimensiunile funcției de repartiție şi a densității de probabilitate. Funcția de repartiție **F(x)**, ca orice probabilitate, este o mărime adimensională. Cât despre densitatea de probabilitate **f(x)**, se vede din formula (2.12) că ea are dimensiunea inversă variabilei aleatoare.

Exemplul de calcul 2.1. Funcția de repartiție a variabilei aleatoare **X** este dată prin :

$$F(x)=\begin{cases}0 & pentru \;\; x\le 0\\ \alpha.x^2 & pentru \;\; 0<x\le 1\\ 1 & pentru \;\; x>1\end{cases}$$

a) Să se găsească coeficientul α.

b) Să se găsească densitatea de probabilitate **f(x).**

c) Să se găsească probabilitatea pentru variabila **X** de a se găsi în intervalul **0,25 – 0,5**.

Rezolvare. a). Funcția de repartiție fiind continuă, pentru **x = 1** avem $\alpha.\mathbf{x}^2 = \mathbf{1}$, de unde $\alpha = \mathbf{1}$.

b) Densitatea de probabilitate a variabilei **X** va fi:

$$f(x)=F'(x)=\begin{cases}0 & pentru \quad x\le 0\\ 2x & pentru \quad 0<x\le 1\\ 0 & pentru \quad x>1\end{cases}$$

c) Conform formulei (2.18) :

$P(0,25<X<0,5) = F(0,5) - F(0,25) = 0,5^2 - 0,25^2 = 0,1875$

Material cu drept de autor

CAP.3. LEGI CLASICE DE PROBABILITATE

3.1. Legea binomială de distribuție

Formarea legii de distribuție binomiale are la bază schema binecunoscută a matematicianului elvețian Nicolas Bernoulli (1687-1759).

Schema, numită schema bilei revenite, constă în extragerea unei bile dintr-o urnă în care se află două tipuri de bile, iar după consacrarea rezultatului extracției, bila se restituie, deci realizarea unui anumit *eveniment* (de exemplu extragerea unei bile de culoare prestabilită) fiind independentă de realizarea celorlalte evenimente.

Pentru realizarea experimentului binomial putem considera existența unei urne cu bile albe (**A**) şi negre (**N**). Se va nota probabilitatea extragerii unei bile albe cu **P(A) = q**, iar probabilitatea extragerii unei bile negre cu **P(N) = p = 1 – q**. Se vor efectua prelevări succesive cu reintroducerea bilelor, înapoi în urnă, după consumarea evenimentului.

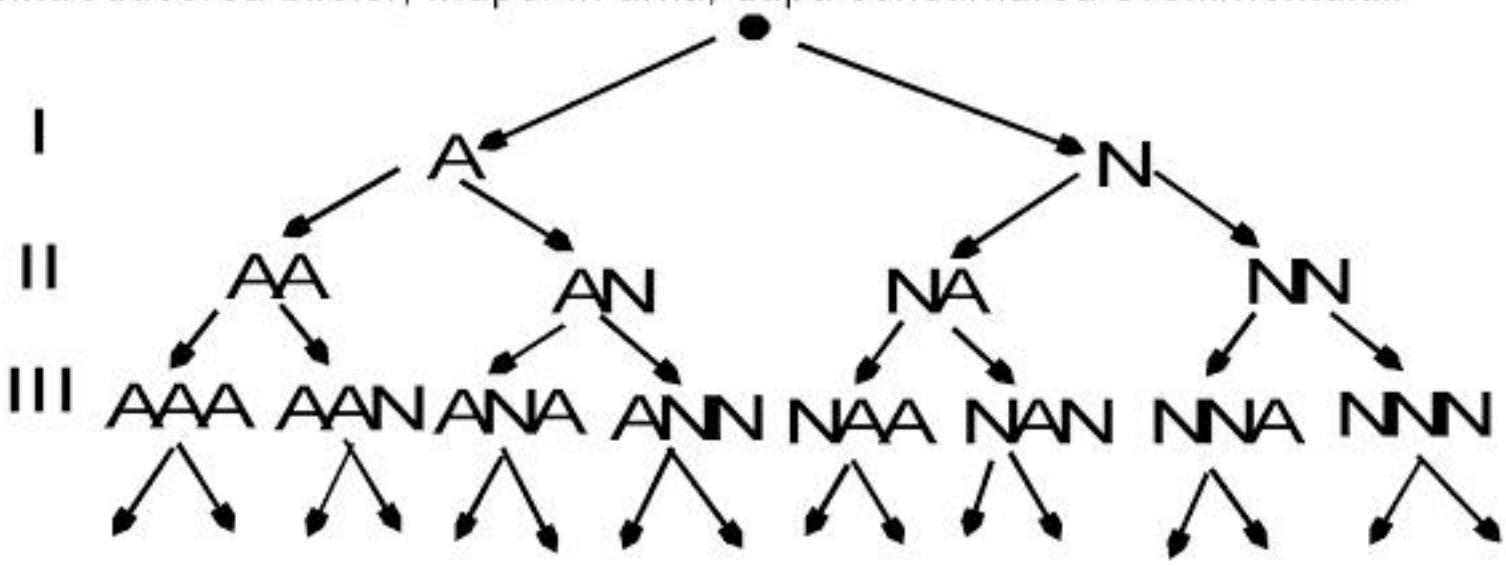


Fig.3.1. Schema arborelui extracției bilelor albe (**A**) sau negre (**N**) din urnă

Urmărind schema din figura 3.1 se poate observa că la extracția I-a se obține numai **A** sau **N**. Dacă în prima extracție rezultatul a fost **A**, în extracția a II-a se obține **AA** sau **AN**, iar dacă rezultatul primei extracții a fost **N**, în cel de-al doilea pas se poate obține **NA** sau **NN**. Întrucât la fiecare extracție sunt posibile două alternative, în general, în **n** extracții numărul alternativelor posibile este $\mathbf{2^n}$.

Procedeul permite să se determine legea de probabilitate a variabilei **X** care corespunde numărului extracțiilor succesive. După cum se observă în figura 3.1, rezultatul celei de a III-a extracții este:

AAA AAN ANA ANN NAA NAN NNA NNN

Valorile şi combinațiile de rezultate la o extracție sunt condiționate de rezultatele la extracția precedentă.

Material cu drept de autor

În a **n**-a extracție, probabilitatea ca să apară de **n** ori evenimentul **A** este $\mathbf{q^n}$ şi probabilitatea ca să apară de **n** ori evenimentul **N** este $\mathbf{p^n}$.

Dacă procedăm la **n** probe consecutive, probabilitatea $\mathbf{P_n(x)}$ pentru ca evenimentul **A** să se realizeze exact de **x** ori şi, firește că, evenimentul **N** să se realizeze de **n – x** ori, va fi, în cazul când nu se specifică ordinea evenimentelor:

$$P_n(x) = P(X = x) = C_n^x p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x . q^{n-x} \qquad (3.1)$$

La rezolvarea problemelor cu factoriale mari ale argumentelor se poate folosi relația aproximativă a lui Steerling:

$$n! \approx \sqrt{2.\pi}\ n^{n+\frac{1}{2}}.e^{-n} \qquad (3.2)$$

sau cu o aproximație mai exactă se poate folosi relația:

$$n! \approx \sqrt{2.\pi}\ n^{n+\frac{1}{2}}.e^{-n+\frac{1}{12.n}} \qquad (3.3)$$

Repartiția de probabilitate a variabilei aleatoare discrete **X**, privind apariția evenimentului **A** este:

$$X : \begin{pmatrix} 0 & 1 & \dots & x & \dots & n \\ q^n & C_n^1 p q^{n-1} & \dots & C_n^x p^x q^{n-x} & \dots & p^n \end{pmatrix} \qquad (3.4)$$

şi poartă denumirea de **lege de repartiție binomială**. Numerele **p** şi **q** sunt determinate de condițiile experienței, iar numărul **n** de probe este fixat dinainte.

Dacă se caută probabilitatea ca evenimentul **A** să se realizeze cel puțin de **x** ori în cursul a **n** probe această probabilitate va fi egală cu suma:

$$p^n + C_n^1 p^{n-1} q + \dots + C_n^x p^{n-x} q^x \qquad (3.5)$$

adică este egală cu suma primilor **x + 1** termeni ai dezvoltării binomului $(p+q)^n$.

Între termenul $\mathbf{P_n(x)}$ şi termenul $\mathbf{P_n(x+1)}$ se formează raportul:

$$\frac{P_n(x+1)}{P_n(x)} = \frac{p}{q} \times \frac{n-x}{x+1} \qquad (3.6)$$

deci, $\mathbf{P_n(x)}$ creşte o dată cu **x**, atât timp cât **x** este inferior sau egal cu **n.p-q**.

Parametrii statistici ai repartiției binomiale $(p + q)^n$ sunt:

Media (speranța matematică):

$$M(X) = \sum_{i=1}^{n} x_i . p_i = m = n.p \qquad (3.7)$$

în care caz probabilitățile sunt date de expresia:

Material cu drept de autor

$$P(x,n) = C_n^x . p^x . q^{n-x}$$
$$(p+q) = 1 \qquad (3.8)$$

Dispersia (varianța):

$$\mathbf{D(X) = \sigma^2 = n.p.q} \qquad (3.9)$$

Abaterea standard:

$$\sigma(X) = \sigma = \sqrt{n.p.q} \qquad (3.10)$$

Momentul centrat de ordinul 3:

$$\mu_3 = \mathbf{n.p.q.(q - p)} \qquad (3.11)$$

Momentul centrat de ordinul 4:

$$\mu_4 = \mathbf{3.n^2.p^2.q^2 + n.p.q.(1 - 6.p.q)} \qquad (3.12)$$

Asimetria:

$$\alpha = \frac{q-p}{\sqrt{n.p.q}} \qquad (3.13)$$

Excesul:

$$E = \frac{1-6.p.q}{n.p.q} \qquad (3.14)$$

Forma repartiției binomiale depinde de valorile **p** și **q** și de valoarea exponentului **n**. Dacă **p = q** repartiția este, evident, simetrică, deoarece termenii echidistanți de la cele două capete ale dezvoltării binomului lui Newton sunt identici. Dacă **p** şi **q** nu sunt egali, creşterea lui **n** micşorează asimetria. Tendința spre simetrie poate fi observată în figura 3.2, în care se dau poligoanele de frecvență ale repartiției binomiale $(\mathbf{0,9 + 0,1})^n$ pentru **n = 10, 50, 100**. Se demonstrează (şi faptul reiese clar din figura 3.2) că repartiția binomială are ca limită repartiția normală, dacă **n** este suficient de mare pentru a face **p – q** mic în comparație cu abaterea standard: $\sqrt{n.p.q}$.

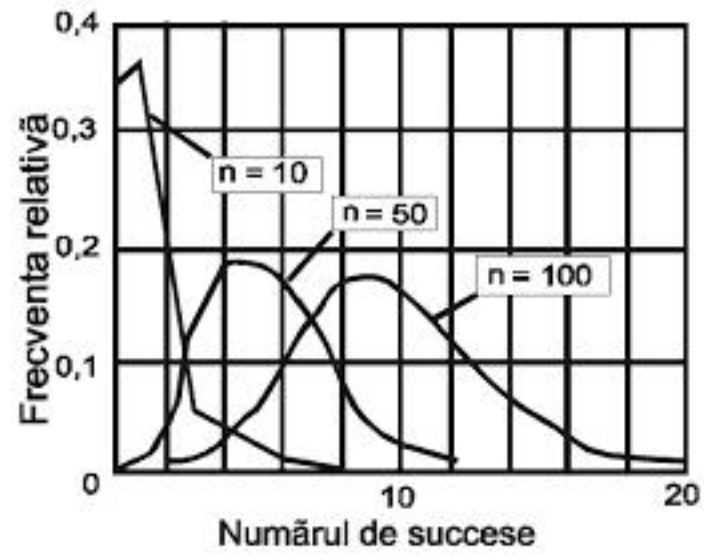


Fig.3.2. Poligoane de frecvență ale repartiției binomiale

Pentru o valoare **p** dată, numărul minim de observații (încercări) **n**, ca repartiția să aibă aspectul de repartiție normală se obține din relația:

$$n = \frac{9}{p.q} \qquad (3.15)$$

Material cu drept de autor

De exemplu, pentru **p = 0,1 (q = 0,9)** vom avea: **n = 9/(0,1x0,9) = 100**.

Probabilitățile **P (n; x)** sunt tabelate, fapt care uşurează operațiunile de calcul în care se utilizează modelul binomial.

Când **n** este mare **P** satisface condiția:

$$\frac{1}{n-1} < P < \frac{n}{n+1} \tag{3.16}$$

şi probabilitățile **P (n; x)** pot fi calculate cu relația aproximativă:

$$P(n;x) = \frac{1}{np(1-p)} f(z) \tag{3.17}$$

unde:

$$z = \frac{x - n.p}{\sqrt{n.p(1-p)}} \tag{3.18}$$

Dacă **n** $\geq$ **50** şi **n.p** $\geq$ **4**, probabilitatea **P(x; n)** poate fi determinată prin relația de aproximare folosind tabelele cu valorile funcției lui Laplace:

$$P(x;n) \approx \Phi\left(\frac{x + \frac{1}{2} - n.p}{\sqrt{n.p(1-p)}}\right) - \Phi\left(\frac{x - \frac{1}{2} - n.p}{\sqrt{n.p(1-p)}}\right) \tag{3.19}$$

În mod similar, probabilitatea căutată pentru variabila considerată să se afle între două limite, se determină cu relația:

$$P(x_1 \leq A \leq x_2) \approx \Phi\left(\frac{x_2 + \frac{1}{2} - n.p}{\sqrt{n.p(1-p)}}\right) - \Phi\left(\frac{x_1 - \frac{1}{2} - n.p}{\sqrt{n.p(1-p)}}\right) \tag{3.20}$$

Exemplu de calcul, 3.1. Se supune recepției un lot de piese de schimb achiziționate de la furnizor din care se cunoaşte că **1%** nu îndeplinesc normele de calitate, lot din care se prelevează întâmplător cinci unități. Se cere să se stabilească valoarea medie **m**, dispersia **D(X)**, cât şi probabilitatea de a identifica **0, 1, 2,...,5** unități de calitate necorespunzătoare.

Rezolvare. Întrucât volumul eşantionului este **n = 5**, iar **p = 0,01**, respectiv **q = 1 – p = 0,99** rezultă: **m = n.p = 0,05** respectiv **D(X) = n.p.q = 0,0495** $\approx$ **0,95**.

Probabilitatea ca între cele cinci piese din eşantion să nu de identifice nici una de calitate necorespunzătoare se determină astfel:

$$\mathbf{P(5;0) = 0,99^5 = 0,95099}$$

Material cu drept de autor

Cu ajutorul relației de recurență (3.6) obținem şi celelalte probabilități. De exemplu, probabilitatea ca din cele cinci unități prelevate să identificăm una necorespunzătoare va fi:

$$P(5;1) = P(5;0)\frac{5-0}{0+1}\cdot\frac{0{,}01}{0{,}99} = 0{,}04803$$

Mai departe valorile obținute descresc rapid, încât putem aprecia că este improbabil ca în cadrul unui eşantion să identificăm mai mult de o unitate necorespunzătoare. Iată valorile acestor probabilități: **P(5;2) = 0,00097; P(5;3) = 0,00001; P(5;4) = 0,00000**.

Exemplu de calcul, nr.3.2. Se supune recepției de calitate un lot de **n = 8000** produse livrabile în loturi colaboratorilor externi şi în care se ştie, din cercetări statistice prealabile, că fracțiunea defectă probabilă este de **0,1%**. Se cere să se determine probabilitățile ca în lotul de **8000** de produse să se depisteze: **a) 10** produse rebut; **b)** nu mai mult de **20** produse rebut.

Rezolvare. Probabilitatea identificării a **10** produse rebut este:

$$P(x=10) = P(10;8000) = C_{8000}^{10}(0{,}001)^{10}(0{,}999)^{7990} =$$

$$= \frac{8000}{10!\,7990!}(0{,}001)^{10}(0{,}999)^{7990}$$

La punctul **b)** se va calcula:

$$P(0 \le x \le 20) = \sum_{x=0}^{20} P(x;8000) = \sum_{x=0}^{20} C_{8000}^{x}(0{,}001)^{x}\ (0{,}999)^{8000-x}$$

Date fiind dificultățile de calcul, se poate utiliza relația aproximativă (3.21), în care:

$$z = \frac{|10 - 8000\times 0{,}001|}{\sqrt{8000.0{,}001.0{,}999}} = \frac{2}{2{,}827} = 0{,}71$$

Din Anexa 1 a valorilor funcției de frecvență a legii normale normate se obține **f(0,71) = 0,3101**, deci răspunsul la punctul **a)** este:

$$P(10;8000) = \frac{0{,}3101}{2{,}827} \approx 0{,}11$$

iar pentru punctul **b)** rezultă:

$$z_1 = \frac{20+\frac{1}{2}-8}{2{,}827} \approx 4{,}20; \quad z_2 = \frac{0-\frac{1}{2}-8}{2{,}827} \approx -3{,}01$$

Din tabelul funcției lui Laplace se identifică Φ**(4,20)** $\approx$ **0,49987**, Φ**(- 301)** = –**0,4987** deci probabilitatea căutată va fi **P = 0,499987 – (-0,4987)**, ceea ce înseamnă că practic este aproape cert că printre cele **8000** de exemplare care formează lotul nu se vor găsi mai mult de **20** de exemplare rebut **P(x>20)**, este probabilitatea evenimentului contrar şi este egală cu: **1 – 0,9987 = 0,0013**.

Material cu drept de autor

Exemplul de calcul 3.3. O centrală telefonică este solicitată zilnic de **n = 10 000** de ori. S-a constatat că probabilitatea de a satisface o cerere este de **0,8**. Care este probabilitatea ca numărul solicitărilor la care s-a răspuns într-o zi să fie cuprins între **7 900** şi **8 100** ?

Rezolvare. Notăm cu α numărul cererilor satisfăcute de centrala respectivă într-o zi. Vom avea:

m = n.p = 10000 x 0,8 = 8000

$$\sigma = \sqrt{n.p.q} = \sqrt{10000\,x\,0{,}8\,x\,0{,}2} = 40$$

Evenimentul:

7 900 < α < 8100

se mai poate scrie:

|α - m| < 100

sau:

$$\left|\frac{\alpha - m}{\sigma}\right| < 2{,}5$$

Deci:

P(7900 < α < 8100) =

$$= P\left(\left|\frac{\alpha - m}{\sigma}\right| < 2{,}5\right) = 2.\Phi(2{,}5) = 2\times 0{,}49379 = 0{,}98758$$

3.2. Legea hipergeometrică de distribuție

Schema de formare a repartiției hipergeometrice se aseamănă cu cea a formării repartiției binomiale. Deosebirea constă în faptul că în timp ce la repartiția binomială, după consumarea rezultatului unei extracții bila respectivă se reintroducea în urnă, în cazul repartiției hipergeometrice nu se mai reintroduce bila respectivă în urnă. Această schemă se mai numeşte şi schema bilei fără revenire.

Pentru demonstrație se consideră că într-o urnă sunt **N** bile dintre care **D** albe şi **G = N-D** negre. Se extrag **n** bile, una câte una, fără revenirea bilei extrase din urnă (sau se scot **n** bile deodată). Se cere să se afle probabilitatea ca din cele **n** bile extrase **d** să fie albe şi **g = n – d** să fie negre.

Presupunem că se iau n bile deodată. Numărul cazurilor posibile este: C_N^n.

Un grup de **d** bile albe, din cele **D** bile albe existente în urnă, poate fi luat în C_D^d moduri; un grup de **g** bile negre, din cele **G** bile negre, poate fi

Material cu drept de autor

luat în C_G^g moduri. Fiecare grup de **d** bile albe poate fi asociat cu fiecare grup de **g** bile negre. Deci numărul cazurilor favorabile va fi: $C_D^d.C_G^g$.

Probabilitatea ca dintre cele **n** bile extrase **d** să fie albe şi **g** să fie negre este:

$$P(d,n;D,N)=\frac{C_D^d C_G^g}{C_N^n} \tag{3.21}$$

Distribuția va fi:

$$X:\begin{pmatrix} 0 & 1 & \dots & x & \dots & n \\ \frac{C_D^0.C_G^n}{C_N^n} & \frac{C_D^1.C_G^{n-1}}{C_N^n} & \dots & \frac{C_D^x.C_G^{n-x}}{C_N^n} & \dots & \frac{C_D^n.C_G^0}{C_N^n} \end{pmatrix} \tag{3.22}$$

Funcția de distribuție este:

$$\sum_{x=0}^{n} P_x=\sum_{x=0}^{n}\frac{C_D^x\times C_G^{n-x}}{C_N^n} \tag{3.23}$$

Parametrii principali ai acestei distribuții sunt:

Abaterea medie:

$$M(X)=m=n.p \tag{3.24}$$

Dispersia:

$$D(X)=\sigma^2=n.p.q.\frac{N-n}{N-1} \tag{3.25}$$

Abaterea medie pătratică:

$$\sigma(X)=\sigma=\sqrt{n.p.q.\frac{N-n}{N-1}} \tag{3.26}$$

în care: **p = d/n; q = g/n; p + q = 1; d + g = n**.

La valori mici ale raportului **n/N**, repartiția hipergeometrică tinde spre repartiția binomială. Atunci când **n ≤ 0,1.N** repartiția hipergeometrică devine deja foarte apropiată de repartiția binomială astfel încât diferența dintre acestea se poate neglija.

Exemplul de calcul 3.4. Un cumpărător se duce la magazin şi cumpără un produs pentru care trebuie să plătească 40.000 lei. El are în buzunar 10 hârtii de câte 10.000 lei dintre care 4 sunt deteriorate. Care este probabilitatea ca în suma plătită la întâmplare să intre 2 hârtii deteriorate ?

Rezolvare. Suntem în cazul schemei bilei neîntoarse (nerevenite): **N = 10; n = 4; D = 4; G = 6; d = 2; g = 2**. Probabilitatea cerută va fi:

$$P(d,n;D,N)=\frac{C_D^d C_G^g}{C_N^n}=\frac{C_4^2.C_6^2}{C_{10}^4}=\frac{3}{7}$$

Material cu drept de autor

Exemplul de calcul 3.5. Într-o excursie au plecat 20 de studenți: 7 din anul I; 7 din anul II şi 6 din anul III. Pe drum au fost aleşi la întâmplare 9 studenți să meargă pe alt traseu. Care este probabilitatea să fie aleşi câte 3 din fiecare an ?

Rezolvare. Suntem în cazul schemei bilei neîntoarse (nerevenite): **N = 20; n = 9; D_1 = 7; D_2 = 7; D_3 = 6; $d_1 = d_2 = d_3 = 3$**. Probabilitatea căutată va fi:

$$P = \frac{C_7^3 . C_7^3 . C_6^3}{C_{20}^9} = \frac{1225}{8398} = 0,146$$

Exemplul de calcul 3.6. pentru controlul de recepție a unui lot de **N** piese se extrag deodată **n** piese (**n < N**). Ştiind că în lot avem **a** piese rebut şi **b** piese bune (**a + b = N**) să se scrie tabloul de distribuție a variabilei aleatoare care reprezintă numărul de piese rebut dintre cele **n** extrase. Să se calculeze valoarea medie şi dispersia acestei variabile.

Rezolvare. Variabila aleatoare **X** poate lua valorile **0, 1,...,a** cu probabilitățile:

$$P(X = k) = \frac{C_a^k . C_b^{n-k}}{C_N^n}$$

deci:

$$X : \begin{pmatrix} k \\ \dfrac{C_a^k . C_b^{n-k}}{C_N^n} \end{pmatrix}, \quad k = 0,1,\ldots,n$$

Avem:

$$M(X) = \frac{1}{C_N^n} \sum_{k=1}^{n} k . C_a^k . C_b^{n-k} = \frac{a}{C_N^n} \sum_{k=1}^{n} C_{a-1}^{k-1} . C_b^{n-k}$$

Dar:

$$\sum_{k=1}^{n} C_{a-1}^{k-1} . C_b^{n-k} = \sum_{k=0}^{n-1} C_{a-1}^{k} . C_b^{n-1-k} = C_{a+b-1}^{n-1} = C_{N-1}^{n-1}$$

şi:

$$M(X) = a \frac{C_{N-1}^{n-1}}{C_N^n} = \frac{n}{N} a \qquad (3.27)$$

Notăm:

p = a/N; q = b/N; deci: **M(X) = n.p;**

$M(X^2)$ = M[X(X – 1)] + M(X)

Dar:

Material cu drept de autor

$$C_N^n.M\left[X(X-1)\right]=\sum_{k=2}^{n}k.(k-1).C_a^k.C_b^{n-k}=$$

$$=a.(a-1)\sum_{k=2}^{n}C_{a-2}^{k-2}.C_b^{n-k}=a.(a-1).C_{a+b-2}^{n-2}=a.(a-1).C_{N-2}^{n-2}$$

de unde:

$$M(X^2)=n\frac{a}{N}(a-1)\frac{n-1}{N-1}+n\frac{a}{N}=\frac{n.a}{N.(N-1)}\left[a.(n-1)+N-n\right]$$

Dispersia va fi:

$$D(X)=\frac{n.a.b}{N^2}\,\frac{N-n}{N-1}=n.p.q\,\frac{N-n}{N-1} \qquad (3.28)$$

3.3. Legea de distribuție a lui Poisson

O posibilitate de realizare a modelului Poisson se poate face pornind de la legea binomială în condițiile în care **p** – probabilitatea de apariție a evenimentului urmărit este mică – iar volumul eşantionului este suficient de mare, astfel încât să aibă loc condiția **n.p** = λ = constant.

În aceste condiții:

$$P(n,x)=C_n^x p^x(1-p)^{n-x} \qquad (3.29)$$

devine la limită ($n\to\infty$ si p $\to 0$) :

$$P_x=\frac{\lambda^x}{x!}e^{-\lambda} \qquad (3.30)$$

Variabila **X** care poate lua valorile **x = 0, 1, 2, 3, ...** cu probabilitățile **P(x)** determinate cu ajutorul relației (3.30) se spune că urmează legea distribuției Poisson. Prin urmare, expresia (3.30) este expresia funcției de frecvență a distribuției Poisson.

Distribuția acestei variabile este:

$$\mathbf{X}:\begin{pmatrix}0 & 1 & 2 & 3 & \dots\\ e^{-\lambda} & \frac{\lambda}{1!}e^{-\lambda} & \frac{\lambda^2}{2!}e^{-\lambda} & \frac{\lambda^3}{3!}e^{-\lambda} & \dots\end{pmatrix} \qquad (3.31)$$

în care λ se numeşte parametrul distribuției.

Funcția de repartiție este:

Material cu drept de autor

$$F(x) = P(X < x) = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} e^{-\lambda} = 1 \qquad (3.32)$$

$$(x = 1,2,...)$$

Repartiția Poisson este determinată numai de un singur parametru şi anume: λ, de aceea ea se tabelează simplu.

Legea lui Poisson este nesimetrică, dar această nesimetrie devine mai mică odată cu creşterea valorii lui λ.

Valoarea medie şi dispersia sunt:

$$M(X) = m = D(X) = \sigma^2 = \lambda \qquad (3.33)$$

Egalitatea valorilor **M(X)** şi **D(X)** nu este în contradicție cu sensul respectării dimensiunilor, deoarece mărimea aleatorie **X** care se supune legii lui Poisson este adimensională.

Abaterea medie pătratică este:

$$\sigma(X) = \sqrt{\lambda} \qquad (3.34)$$

Momentul centrat de ordinul 3:

$$\mu_3 = \lambda \qquad (3.35)$$

Momentul centrat de ordinul 4:

$$\mu_4 = \lambda + 3.\lambda^2 \qquad (3.36)$$

Asimetria este:

$$\alpha = \frac{1}{\sqrt{\lambda}} \qquad (3.37)$$

Excesul este:

$$E = \frac{1}{\lambda} \qquad (3.38)$$

După cum se observă, asimetria şi excesul sunt totdeauna pozitive, adică $\alpha > 0$ şi $E > 0$.

Schema lui Poisson este foarte utilă în aplicații. Ea este folosită la controlul statistic al produselor industriale atunci când probabilitatea obținerii unei piese defecte este foarte mică. De asemenea, se mai aplică în probleme de management, de cercetări operaționale, telefonie, în unele probleme de fizică etc.

Scriindu-se raportul între două valori consecutive ale funcției de frecvență se obține:

$$\frac{P_{x+1}}{P_x} = \frac{\dfrac{\lambda^{x+1} e^{-\lambda}}{(x+1)!}}{\dfrac{\lambda^x e^{-\lambda}}{x!}} = \frac{\lambda}{x+1} \qquad (3.39)$$

Material cu drept de autor

de unde se obține următoarea formulă de recurență foarte utilă pentru calculul funcției de frecvență:

$$P_{x+1} = P_x \frac{\lambda}{x+1} \tag{3.40}$$

Pentru **x = 0** va fi:

$$P_{x=0} = e^{-\lambda} \tag{3.41}$$

iar celelalte valori ale funcției de frecvență se obțin din aproape în aproape utilizându-se relația (3.40).

Suma tuturor probabilităților $\mathbf{P_x = f(x)}$ obținute atunci când **x** ia valorile **0, 1, 2, 3, ...** este egală cu unitatea:

$$\sum_{x=0}^{+\infty} \frac{\lambda^x e^{-\lambda}}{x!} = 1 \tag{3.42}$$

Exemplu de calcul 3.7. Se consideră aplicarea modelului evenimentelor rare pentru recepția unor loturi de câte **5000** de exemplare (**n = 5000)**. Din analiza desfăşurării procesului de fabricație se cunoaşte că fracțiunea defectă probabilă este de **0,2% (p = 0,002)**. Care este probabilitatea ca în lot să se găsească:

a) 15 exemplare necorespunzătoare;

b) mai mult de 10 exemplare necorespunzătoare.

Rezolvare: Întrucât **n = 5000** este suficient de mare, iar **p = 0,002** este suficient de mic, putem aproxima probabilitatea:

$$P(15;5000) = C_{5000}^{15}(0{,}002)^{15}(0{,}998)^{4985}$$

cu distribuția Poisson de parametru $\lambda = n.p = 500.0{,}002 = 10$.

$$f(15;10) = e^{-10}\frac{10^{15}}{15!} = 0{,}0347$$

Pentru punctul b) se determină mai întâi probabilitatea evenimentului opus (probabilitatea ca în lot să se găsească cel mult 10 exemplare defecte):

$$P(0 \le x \le 10) = P(0 \le x \le 10;5000) = \sum_{d=0}^{10} P(d;5000) \approx \sum_{d=0}^{10} f(d;10) = 0{,}583$$

Atunci probabilitatea ca în lot să se găsească mai mult de 10 exemplare defecte va fi:

$$P(X > 10) = 1 - 0{,}583 = 0{,}417$$

3.4. Legea normală de distribuție

Legea de distribuție normală, numită şi legea lui Gauss, joacă un rol deosebit de important în teoria probabilităților. Particularitatea fundamentală care o deosebeşte de celelalte legi constă în aceea că ea este o **lege limită**

Material cu drept de autor

spre care tind celelalte legi, în condițiile ce se întâlnesc frecvent în aplicațiile practice.

Studiată de Gauss la sfârşitul secolului al XVIII-lea, legea a fost aplicată de Laplace la erorile observațiilor în cazul repetării încercărilor.

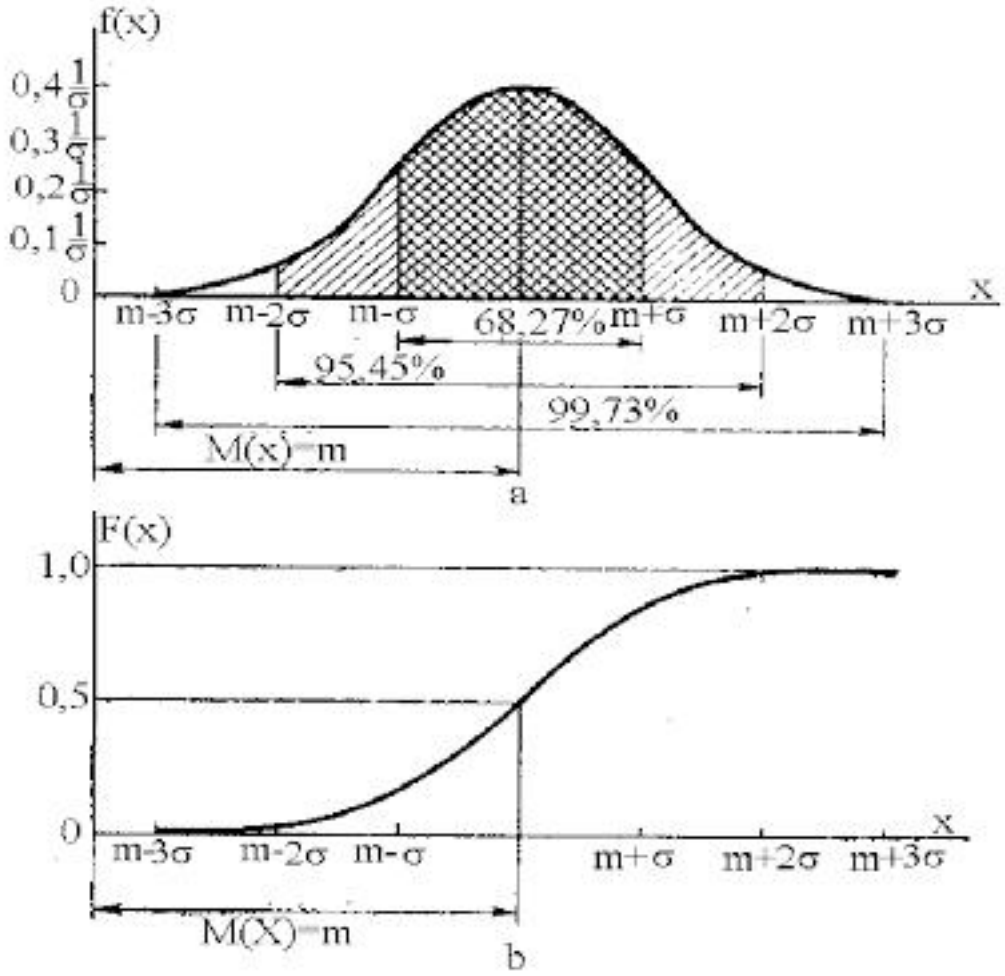


Fig.3.3. Densitate de probabilitate (a) şi funcția de repartiție (b) pentru distribuția normală

Se poate arăta că suma unui număr suficient de mare de variabile independente (sau slab legate), care urmează o lege oarecare de repartiție, tinde spre o lege normală. Majoritatea variabilelor aleatoare întâlnite în practică (erori de măsură, erori de tir etc.), pot fi considerate ca sume de un număr important de termeni, erori elementare, datorate fiecare unei cauze independente de celelalte. Oricare ar fi legile erorilor elementare, particularitățile repartiției lor nu apar în suma unui număr mare de aceste erori. Singura limitare impusă este ca fiecare dintre aceste erori să joace în sumă un rol relativ puțin important. Dacă una dintre erorile aliatoare prevalează net asupra celorlalte, aceasta determină legea de repartiție a sumei.

Densitatea probabilității variabilei aleatoare care are o repartiție normală depinde de doi parametri: **m = M(X)** şi **σ(X) =** $\sqrt{D(X)}$ fiind generată de expresia:

$$\int_{-\infty}^{+\infty} f(x).dx = 1$$

$$f(x) = \frac{1}{\sigma\sqrt{2\cdot\pi}} e^{-\frac{(x-m)^2}{2\cdot\sigma^2}} \qquad (3.43)$$

în care **x** ia valori între -∞ şi +∞.

Funcția de repartiție este definită de expresia:

Material cu drept de autor

$$F(x)=\int_{-\infty}^{x} f(x)dx=\frac{1}{\sigma.\sqrt{2.\pi}}\int_{-\infty}^{x} e^{-\frac{(x-m)^2}{\sigma^2}}.dx \qquad (3.44)$$

Pentru a demonstra că funcția **f(x)** este o densitate de repartiție trebuie să arătăm că:

Folosim integrala Euler – Poisson:

$$\int_{0}^{\infty} e^{-y^2}.dy=\frac{\sqrt{\pi}}{2}$$

Facem schimbarea de variabilă:

$$\frac{x-m}{\sigma\sqrt{2}}=y\ ,\ x=m+\sigma.y.\sqrt{2}\ ,\ dx=\sigma.\sqrt{2}.dy$$

Rezultă:

$$\frac{1}{\sigma.\sqrt{2.\pi}}\int_{-\infty}^{+\infty} e^{-\frac{(x-m)^2}{2.\sigma^2}}.dx=\frac{1}{\sigma.\sqrt{2.\pi}}\int_{-\infty}^{+\infty} e^{-y^2}.\sigma.\sqrt{2}.dy=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{+\infty} e^{-y^2}.dy$$

Întrucât e^{-y^2} este o funcție simetrică de y, se poate scrie:

$$\int_{-\infty}^{+\infty} e^{-y^2}.dy=2.\int_{0}^{\infty} e^{-y^2}.dy=2\frac{\sqrt{\pi}}{2}=\sqrt{\pi}$$

Deci:

$$\frac{1}{\sqrt{\pi}}\int_{-\infty}^{+\infty} e^{-y^2}.dy=1$$

Graficul funcției **f(x)** este simetric față de dreapta **x = m**. Într-adevăr:

$$f(m+\alpha)=\frac{1}{\sigma\sqrt{2.\pi}}.e^{-\frac{(m+\alpha-m)^2}{2.\sigma^2}}=\frac{1}{\sigma.\sqrt{2.\pi}}e^{-\frac{\alpha^2}{2.\sigma^2}}$$

$$f(m-\alpha)=\frac{1}{\sigma\sqrt{2.\pi}}.e^{-\frac{(m-\alpha-m)^2}{2.\sigma^2}}=\frac{1}{\sigma.\sqrt{2.\pi}}e^{-\frac{\alpha^2}{2.\sigma^2}}$$

Se poate constata că în punctul **x = m**, funcția **f(x,m,σ)** are un maxim şi că punctele **x = m** - σ şi **x = m** + σ sunt puncte de inflexiune.

Valoarea maximă a funcției este:

Material cu drept de autor

$$f(m) = \frac{1}{\sigma.\sqrt{2\pi}} = \frac{1}{\sigma} 0{,}3989$$

Funcția de repartiție **F(x) = P(X<x)** satisface condițiile:

$$\mathbf{F(x) > 0} \tag{3.45}$$

$$\mathbf{F(x_1) < F(x_2)} \text{ pentru } \mathbf{x_1 < x_2} \tag{3.46}$$

$$F(+\infty) = P(X < +\infty) = \frac{1}{\sigma\sqrt{2\cdot\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-m)^2}{2\cdot\sigma^2}} dx = 1 \tag{3.47}$$

Indiferent de valorile parametrilor **m** şi σ, graficele funcțiilor **f(x)** au formă de clopot. Parametrul **m** defineşte poziția axei de simetrie, figura 3.4, iar σ stabileşte gradul de "boltire" al graficului, figura 3.5.

Fig.3.4. Comparația curbelor de densitate a probabilității pentru repartițiile Gauss la diferite valori ale parametrului **m** şi aceeaşi dispersie

Momentele simple (inițiale) de ordin **k** se determină cu expresia:

$$\mu_k = \int_{-\infty}^{+\infty} x^k . f(x).dx \tag{3.48}$$

Momentul centrat de ordin k se va determina în felul următor:

$$M_k = \int_{-\infty}^{+\infty} (x-m)^k . f(x;m;\sigma).dx = \frac{1}{\sigma\sqrt{2.\pi}} \int_{-\infty}^{+\infty} (x-m)^k e^{-\frac{(x-m)^2}{2.\sigma^2}} .dx =$$

$$= \frac{\left(\sigma\sqrt{2}\right)^k}{\sqrt{\pi}} \int_{-\infty}^{+\infty} t^k .e^{-t^2} dt; \qquad t = \frac{x-m}{\sigma\sqrt{2}}$$

Integrând prin părți se obține:

$$M_k = \frac{\left(\sigma\sqrt{2}\right)^k}{\sqrt{\pi}} \left[-\frac{1}{2} e^{-t^2} .t^{k-1} \Big|_{-\infty}^{+\infty} + \frac{k-1}{2} \int_{-\infty}^{+\infty} t^{k-2} .e^{-t^2} .dt \right] =$$

$$= \frac{(k-1)\left(\sigma\sqrt{2}\right)^k}{2\sqrt{\pi}} \int_{-\infty}^{+\infty} t^{k-2} .e^{-t^2} .dt$$

de unde:

Material cu drept de autor

$$M_k = (k-1).\sigma^2 M_{k-2} \qquad (3.49)$$

Din expresia (3.49), de recurență, se obține: $M_2 = \sigma^2$; $M_4 = 3.\sigma^4$; $M_6 = 15.\sigma^6$,..., $M_k = (k-1)!!.\ \sigma^2$ unde s-a notat prin **(k – 1)!!** Produsul tuturor numerelor impare.

Valoarea medie se determină în felul următor:

Făcând schimbarea de variabilă: $t = \dfrac{x-m}{\sigma\sqrt{2}}$ se obține:

$$M(X) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} \left(\sigma\sqrt{2}\,t + m\right) e^{-t^2}.dt = \frac{\sigma\sqrt{2}}{\sqrt{\pi}} \int_{-\infty}^{+\infty} t.e^{-t^2}.dt + \frac{m}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-t^2}.dt$$

Se observă că prima integrală este nulă, iar cea de a doua este integrala Euler-Poisson:

$$\int_{-\infty}^{+\infty} e^{-t^2}.dt = 2.\int_{0}^{\infty} e^{-t^2}.dt = \sqrt{\pi}$$

Rezultă:

$$\mathbf{M(X) = m = \mu_1} \qquad (3.50)$$

Deci parametrul **m** este tocmai valoarea medie a lui **X**.

Pentru dispersie:

$$D(X) = \frac{1}{\sigma\sqrt{2.\pi}} \int_{-\infty}^{+\infty} (x-m)^2.e^{-\frac{(x-m)^2}{2.\sigma^2}}.dx$$

Cu aceeaşi schimbare de variabilă: $t = \dfrac{x-m}{\sigma\sqrt{2}}$ se obține:

$$D(X) = \frac{2.\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{+\infty} t^2.e^{-t^2}.dt = \frac{\sigma^2}{\sqrt{\pi}} \left[-t.e^{-t^2}\Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-t^2}.dt \right]$$

de unde:

$$\mathbf{D(X) = \sigma^2 = M_2} \qquad (3.51)$$

Mediana (Me) pentru o variabilă aleatoare repartizată normal este numărul egal cu jumătate din lungimea unui interval de pe axa absciselor simetric în raport cu punctul **m** şi care este baza figurii de arie egală cu jumătate din aria mărginită de axa **Ox** şi curba de repartiție.

Dacă **X** este variabila aleatoare repartizată normal, din definiție rezultă că:

Material cu drept de autor

$$P\left(|X - m| < Me\right) = \frac{1}{2} \tag{3.52}$$

deci este valabilă şi relația:

$$P\left(|X - m| > Me\right) = \frac{1}{2}$$

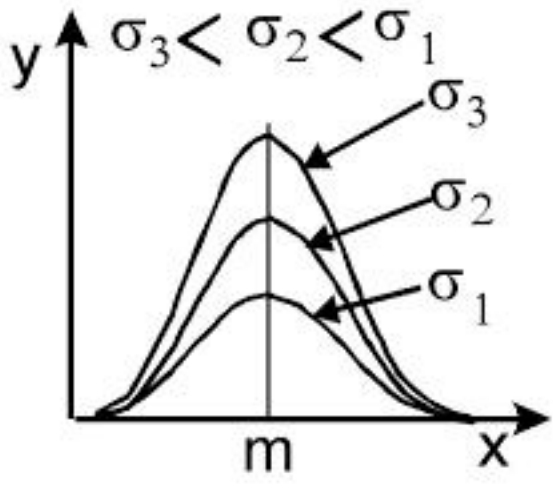


Fig.3.5. Comparația curbelor pentru densitatea de probabilitate a repartiției Gauss pentru diferite valori ale parametrului σ cu valoare medie **m** constantă

Din expresia (3.52), ținând seama de simetria domeniului în raport cu centrul de dispersie rezultă:

$$P\left(|X - m| < Me\right) = 2.\Phi^*\left(\frac{Me}{\sigma}\right) - 1 = \frac{1}{2}$$

de unde:

$$\Phi^*\left(\frac{Me}{\sigma}\right) = 0{,}75$$

iar din tabele rezultă:

$$\frac{Me}{\sigma} = 0{,}674, \;\Rightarrow\; Me = 0{,}674.\sigma \tag{3.53}$$

Deci, cunoscând abaterea medie pătratică σ se poate determina imediat valoarea medianei **Me**.

Dacă pentru caracterizarea dispersiei se utilizează mediana, densitatea de probabilitate a repartiției normale va fi:

$$f(x;m;\sigma) = \frac{\rho}{E\sqrt{\pi}} e^{-\frac{\rho^2}{Me^2}(x-m)^2} \tag{3.54}$$

unde **Me** = $\rho.\sqrt{2}\,\sigma$.

Moda Mo(X) este valoarea variabilei aleatorii care are densitatea maximă de probabilitate. Moda se mai numeşte valoarea cea mai probabilă. Dacă curba de repartiție are un maxim, atunci valoarea mărimii, care corespunde acestui maxim este şi modă. O asemenea curbă se numeşte unimodală sau monomodală. Dacă curba de repartiție are câteva maxime, atunci moda este valoarea care corespunde maximului cel mai mare.

Pentru curba normală teoretică:

Mo = Me = m (3.55)

Material cu drept de autor

3.5. Legea normală normată a lui Laplace

În cele mai multe cazuri practice repartiția normală nu se utilizează sub forma sa inițială definită de densitatea de probabilitate (3.43) sau legea de repartiție (3.44), ci în forma normată. Normarea repartiției constă în trecerea de la mărimea aleatorie **X** la funcția liniară auxiliară, definită de schimbarea de variabilă:

$$z = \frac{x - m}{\sigma}; \quad \mathbf{x = m + \sigma .z;\ dx = \sigma .dz,} \qquad (3.56)$$

pentru care **M(Z) = 0; D(Z) = 1**.

Densitatea de probabilitate a mărimii aleatorii normate **z**, se supune legii lui Gauss şi are următoarea expresie:

$$f(z) = \frac{1}{\sqrt{2.\pi}} e^{-\frac{z^2}{2}} \qquad (3.57)$$

Valorile densității de probabilitate obținute cu expresia (3.57) pot fi tabelate.

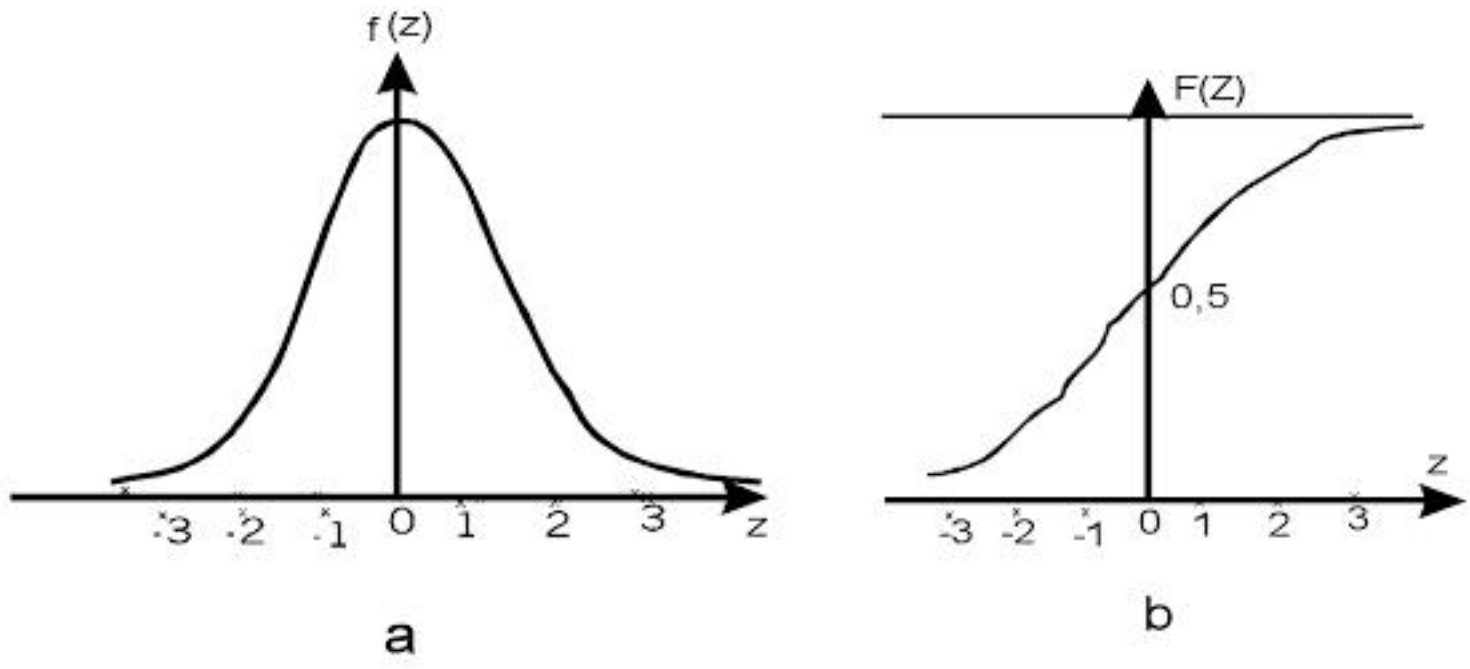


Fig.3.6. Reprezentarea grafică a densității de probabilitate (a) şi a funcției de repartiție (b) pentru legea normală normată

Funcția de repartiție este:

$$F(z) = \frac{1}{\sqrt{2 \cdot \pi}} \int_{-\infty}^{z} e^{-\frac{z^2}{2}} dz \qquad (3.58)$$

Media şi dispersia sunt:

Material cu drept de autor

$$\mathbf{M}[Z] = \mathbf{0};\ \mathbf{D}[Z] = \mathbf{1} \tag{3.59}$$

În figura 3.6 se prezintă trasarea grafică a densității de probabilitate şi a funcției de repartiție pentru distribuția normală normată.

Din egalitățile:

$$1 = \int_{-\infty}^{+\infty} f(z).dz = 2.\int_{0}^{+\infty} f(z).dz = 2\int_{-\infty}^{0} f(z).dz \tag{3.60}$$

rezultă:

$$\int_{0}^{+\infty} f(z).dz = \int_{-\infty}^{0} f(z).dz = \frac{1}{2} \tag{3.61}$$

Pe baza relației (3.61) şi a reprezentării grafice din figura 3.6 rezultă că distribuția normală normată este simetrică față de ordonata maximă corespunzătoare originii.

Ordonata maximă este:

$$f(0) = \frac{1}{\sqrt{2\cdot\pi}} e^{-\frac{0}{2}} = \frac{1}{\sqrt{2\cdot\pi}} = 0{,}3989 \tag{3.62}$$

Derivata a doua:

$$f''(z) = \left(z^2 - 1\right) f(z)$$

se anulează pentru **z = 1** şi **z = -1**. Deci curba densității de probabilitate normale normate are punctele de inflexiune situate la dreapta şi la stânga originii la o distanță egală cu ±**1**.

Cunoscând abaterea medie pătratică σ şi valoarea medie **m** pentru distribuția normală se poate trece uşor de la distribuția normală a variabilei **x** la distribuția normală normată **z** şi invers:

$$f(x) = \frac{1}{\sigma} f(z);\quad z = \frac{x - m}{\sigma} \quad \text{sau} \quad x = z\sigma + m \tag{3.63}$$

Funcția de repartiție **F(x)** a distribuției normale cu media **m** şi abaterea mediei pătratice σ şi funcția de repartiție **F(z)** a distribuției normale normate corespunzătoare au expresiile:

$$F(x) = \frac{1}{\sigma\sqrt{2\cdot\pi}} \int_{-\infty}^{x} e^{-\frac{(x-m)^2}{2\cdot\sigma^2}} dx \tag{3.64}$$

şi respectiv:

$$F(z) = \frac{1}{\sqrt{2\cdot\pi}} \int_{-\infty}^{\frac{x-m}{\sigma}} e^{-\frac{z^2}{2}} dz \tag{3.65}$$

Material cu drept de autor

Dar, din relațiile (3.64) şi (3.65) rezultă:

$$F(z) = F\left(\frac{x-m}{\sigma}\right) \tag{3.66}$$

Relația (3.66) ne permite să transformăm curba funcției de distribuție **F(x)** în curba funcției de distribuție normală normată **F(z)**. Potrivit acestei relații, ordonata unui punct oarecare de pe curba **F(x)** este egală cu ordonata punctului corespunzător de pe curba **F(z)**, iar abscisa punctului respectiv se obține utilizând relația:

$$x = z \cdot \sigma + m \tag{3.67}$$

În practică, de cele mai multe ori, se dă probabilitatea **P** şi se cere să se determine valoarea corespunzătoare a variabilei întâmplătoare. Aceasta înseamnă că se dă valoarea funcției de distribuție şi se cere valoarea variabilei care-i corespunde. Aceste valori ale variabilei se numesc quantili. Quantilul z_p se determină din relația:

$$F(z_p) = P \text{, unde } 0 \le P \le 1 \tag{3.68}$$

De aici necesitatea de a prezenta funcția de distribuție sub forma tabelară. Ținându-se seama de relațiile (3.59) şi (3.60), precum şi de faptul că distribuția normală normată este simetrică față de axa ordonatelor, funcția de repartiție se poate prezenta sub forma:

$$F(z) = \frac{1}{2} + \frac{1}{\sqrt{2\cdot\pi}}\int_0^z e^{-\frac{z^2}{2}}dz \tag{3.69}$$

Integrala definită:

$$\Phi(z) = \frac{1}{\sqrt{2\cdot\pi}}\int_0^z e^{-\frac{z^2}{2}}dz \tag{3.70}$$

reprezentă aria mărginită de curba **f(z)** şi abscisă în intervalul **[0,z]**, adică aria haşurată din figura 3.7 şi se numeşte **funcția integrală Laplace** sau simplu **funcția lui Laplace**.

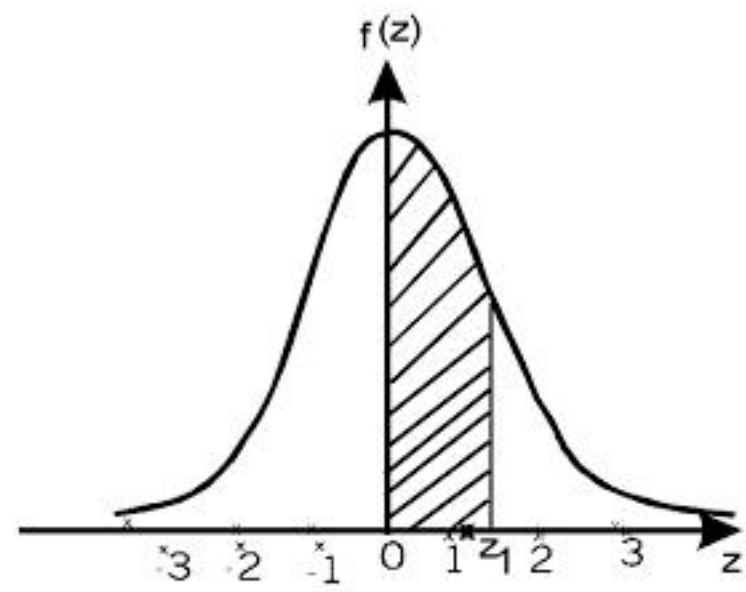


Fig.3.7. Reprezentarea grafică a funcției lui Laplace

În anexă sunt date valorile funcției lui Laplace pentru valorile lui **z** cuprinse între **0** şi **5** exprimate în sutimi.

Aşadar funcția de distribuție **F(z)** se poate scrie sub forma:

$$F(z) = \frac{1}{2} + \Phi(z) \tag{3.71}$$

Material cu drept de autor

În baza acestui rezultat, quantilul z_p se determină din relația:

$$\Phi(z_p) = F(z_p) - \frac{1}{2} = P - \frac{1}{2} \qquad (3.72)$$

adică din probabilitatea dată se scade **1/2**; valorii $\Phi(z)$ astfel obținută îi corespunde în tabelul funcției lui Laplace valoarea z_p, adică quantilul căutat.

De remarcat că:

$$\Phi(0) = 0;\ \Phi(-z) = -\Phi(z);\ \Phi(+\infty) = \frac{1}{2};\ \ \Phi(-\infty) = -\frac{1}{2} \qquad (3.73)$$

Rezultă că aria mărginită de curba **f(z)** şi abscisă în intervalul **(-z,0)** este egală şi de semn contrar cu aria limită de curba **f(z)** în intervalul **(0,z)**.

Pentru caracteristicile fenomenelor tehnico-economice şi sociale se stabilesc în general două limite, aşa încât se pune problema determinării probabilității ca valorile variabilei **X** să se găsească între două valori date x_1 şi x_2, adică:

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x).dx = F(x_2) - F(x_1) \qquad (3.74)$$

Efectuând normarea: $z_1 = (x_1 - m)/\sigma$ şi $z_2 = (x_2 - m)/\sigma$, rezultă:

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x).dx = \int_{z_1}^{z_2} f(z).dz = F\left(\frac{x_2 - m}{\sigma}\right) - F\left(\frac{x_1 - m}{\sigma}\right) = F(z_2) - F(z_1)$$

Dacă x_1 şi x_2 sunt simetrice față de valoarea medie **x = m**, diferențele $x_1 - m$ şi $x_2 - m$ sunt egale şi de semne contrarii, adică:

$$x_1 - m = -(x_2 - m)$$

şi deci:

$$\frac{x_1 - m}{\sigma} = -\frac{x_2 - m}{\sigma} \quad \text{sau } z_1 = -z_2$$

astfel că avem:

$$P(x_1 < X < x_2) = 2 \cdot \Phi(z_2) \qquad (3.75)$$

Rezultă că în multe situații se utilizează dublul funcțiile lui Laplace:

$$2 \cdot \Phi(z) = \frac{2}{\sqrt{2 \cdot \pi}} \int_0^z e^{-\frac{z^2}{2}} dz = \frac{1}{\sqrt{2 \cdot \pi}} \int_{-z}^{+z} e^{-\frac{z^2}{2}} dz$$

fapt pentru care în tabelul 3.3 este tabelată şi dublul funcției lui Laplace.

Exemplu de calcul 3.8. Într-o unitate economică s-au instalat **2000** de becuri noi. Viața medie a unui bec este garantată de producător de **1000** ore de funcționare, cu o abatere medie pătratică de **100** ore. Se cere să se determine:

Material cu drept de autor

1. numărul de becuri care probabil se vor arde în primele **700** de ore de funcționare;

2. numărul de becuri care probabil se vor arde între **900** şi **1300** ore de funcționare,

3. numărul de ore după care probabil se vor arde **10%** din becuri.

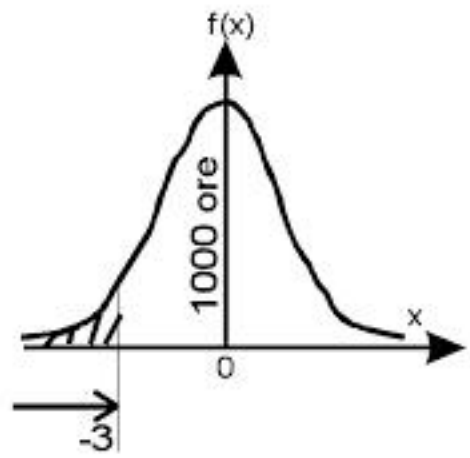


Fig.3.8. Graficul aferent studiului exemplului de calcul 3.8

Rezolvare:

1. Notăm cu **x** durata de funcționare a unui bec. Variabila normală normată va fi:

$$z = \frac{x - m}{\sigma} = \frac{700 - 1000}{100} = -3$$

Aria de sub curba normală, de la $-\infty$ la **z = - 3**, din figura 3.8, furnizează numărul probabil de becuri care au viața sub **700** ore. Din considerentele de simetrie, această arie este aceeaşi cu cea de sub curba normală normată între **z = +3** şi $+\infty$.

Probabilitatea corespunzătoare valorii **z = - 3** este

$$\alpha = 0{,}5 - \Phi(-3) = 0{,}5 - \Phi(3) = 0{,}5 - 0{,}4967 = 0{,}00130$$

Deoarece s-au montat **2000** becuri, numărul probabil de căderi în primele **700** ore este:

$$\mathbf{(2000)x(0{,}00130) \approx 3\,becuri}$$

2. Procedând ca la punctul precedent, se găseşte:

$$z = \frac{900 - 1000}{100} = -1{,}0$$

Din tabelul funcției lui Laplace $\Phi(-1) = 0{,}341$, probabilitatea ca un bec să cadă în primele **900** de ore este **0,5-0,341 = 0,159**, adică numărul probabil de becuri care cad în acest interval de timp este:

$$\mathbf{(2000)x(0{,}159) \approx 318\,becuri}$$

În mod similar, variabila normală normată care corespunde valorii **x = 1300** ore este:

$$\mathbf{x = (1300 - 1000)/100 = 3{,}00}$$

Aria de sub curba normală, de la z = $-\infty$ la **z = +3,0** este egală cu **0,5 +** Φ **(z) = 0,5 + 0,4987 = 0,9987**.

Deci, numărul probabil de becuri care cad între **x = 0** ore şi **x = 1300** ore de funcționare este:

(2000) x (0,9987) $\approx$ **1997** becuri

Material cu drept de autor

Rezultă că numărul probabil de căderi între **x = 900** ore şi **1300** ore este: **1997 – 318 = 1679** becuri

Altfel spus, dacă **318** căderi au avut loc în primele **900** ore, probabil că altele **1679** vor cădea în următoarele **400** ore.

3. Probabilitatea căutată este egală cu **1 – 0,10 = 0,90** căreia îi corespunde o valoare tabelară **z = 1,28**. se poate scrie deci că:

$$-1{,}28 = \frac{x - 1000}{100}$$

de unde, durata de timp căutată va fi egală cu **x = 872** ore. Adică, probabil că **10%** din becuri se vor "arde" în primele **872** ore de funcționare. Rezultă că fiabilitatea becurilor, pentru primele **872** ore de funcționare este de **90%**; fiabilitatea pentru **700** ore de funcționare este:

(1-0,00130) x (100) = 99,86%

fiabilitatea pentru **900** ore de funcționare este:

(1 – 0,159) x (100) = 84,1%

iar fiabilitatea pentru **1300** ore de funcționare este de numai **6,7%.**

3.6. Verificarea corespondenței dintre repartițiile teoretice şi cele empirice

O problemă deosebit de importantă în prelucrarea statistică a datelor experimentale este aceea de a stabili în ce măsură repartiția empirică studiată se apropie sau coincide cu cea normală. În felul acesta se ajunge la problema momentelor prin intermediul cărora funcția de distribuție **F(x)** a unei variabile aleatoare **X** este unic determinată.

Abaterea curbei normale empirice de la cea teoretică se concretizează în abateri de formă – considerate ca asimetrie de la ordonata **x = m** sau excesul sau boltirea caracterizat de existența unei înălțimi mai mari sau mai mici a curbei de frecvențe în raport cu cea teoretică, figura 3.9.

Abaterea de la forma teoretică se apreciază pe baza momentelor. Termenul de moment este împrumutat din mecanică unde este folosit pentru a nota capacitatea forței de a provoca o mişcare. În statistică momentul este folosit în acelaşi sens, forțele fiind înlocuite prin frecvențele absolute ale intervalelor seriei de date statistice. Astfel, dacă vom considera o variabilă **X** a cărei repartiție este:

$$X\begin{pmatrix} x_1, x_2, \ldots x_m \\ n_1, n_2, \ldots, n_m \end{pmatrix}$$

prin definiție momentul de ordinul **k** al repartiției **X** este media aritmetică a puterilor de ordin **k** al abaterilor (x_i-**a**) unde **a** este o constantă aleasă arbitrar. În raport cu valoarea aleasă ca origine, deosebim momente inițiale (când **a = 0**) momente centrate (când **a = m)** şi momente obişnuite (când **a** $\neq$ **0** $\neq$ **m**). Dacă **a = 0** avem momentul inițial de ordinul **k** sau media de

Material cu drept de autor

putere $\mu_0 = 1; \mu_1 = m$ sau $\overline{x}$ (media aritmetică) ş.a.m.d. Dacă **a = m** sau $\overline{x}$, avem momentul centrat de ordinul **k**, $\mu_1 = 0$, $\mu_2 = \sigma^2$.

$$\mu_k = \frac{\sum (x_i - a)^k \cdot n_i}{\sum n_i} \qquad (3.76)$$

Momentele centrate (le vom nota cu **M** pentru a le deosebi de cele inițiale) şi se determină cu relația de recurență (3.49):

$$M_k = (k-1).\sigma^2 . M_{k-2} \qquad (3.77)$$

După cum se poate observa, momentul simplu (inițial) de ordinul **1** reprezintă media aritmetică, iar momentul centrat de ordinul **2** reprezintă abaterea medie pătratică.

Pentru a se putea compara între ele serii statistice diferite, s-a recurs la un moment adimensional, deci care nu este legat de unitatea de măsură a fiecărei serii de date statistice, fiind o valoare abstractă, un raport:

$$\alpha_k = \frac{M_k}{s^k} \qquad (3.78)$$

Momentul α de ordinul **k**, respectiv α_k este raportul dintre momentul centrat de ordinul **k** şi puterea **k** a abaterii standard.

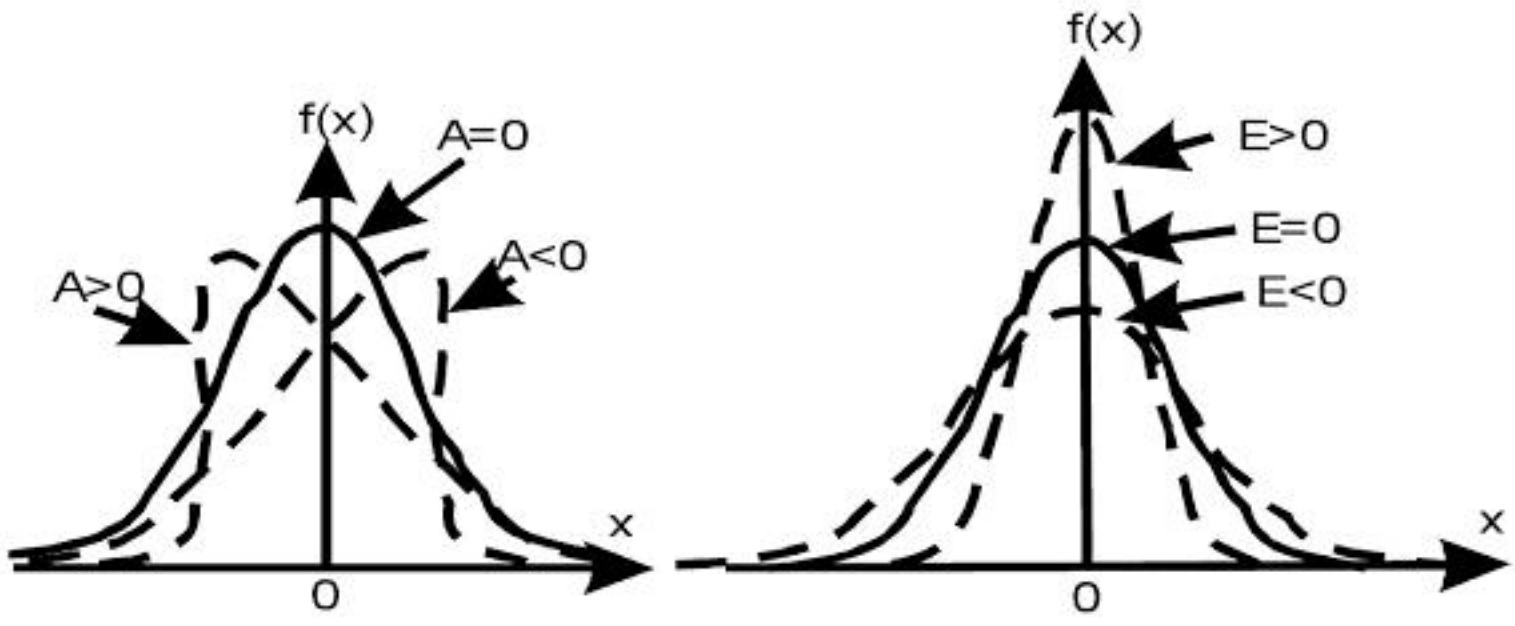


Fig.3.9. Asimetria (a) şi excesul (b) ca abateri de formă de la distribuția normală teoretică

Asimetria. Momentele centrate de ordin impar sunt sumele puterilor impare ale abaterilor față de media aritmetică. Ele vor fi negative pentru abateri negative şi pozitive pentru abateri pozitive. Într-o serie de valori simetrice, cele două grupe de valori pozitive şi negative se compensează şi astfel momentele impare sunt nule. În repartițiile asimetrice vor predomina fie valorile pozitive fie cele negative. Reiese că momentele

Material cu drept de autor

impare pot fi luate ca măsură a asimetriei unui şir (serie de date statistice). Se foloseşte drept coeficient de asimetrie momentul α de ordinul **3**, respectiv α_3. În cazul curbei normale

$$A = \alpha_3 = \mathbf{0} \tag{3.79}$$

Excesul (boltirea). dacă se observă cu atenție expresiile momentelor centrate, se poate constata că momentele centrate de ordin par, ca şi dispersia sau abaterea medie pătratică, sunt o măsură a împrăştierii valorilor din seria statistică. Ca atare, momentele centrate de putere pară mai mare decât **2** scot mai bine în evidență abaterile extreme decât dispersia. De aceea, se folosesc pentru aprecierea excesului (boltirii), adică a gradului de turtire a curbei de repartiție. Termenul de comparație este boltirea curbei empirice. Excesul unei curbe empirice de repartiție se apreciază cu **coeficientul de exces**:

$$E = \alpha_4 - 3 \tag{3.80}$$

În cazul repartiției normale $\alpha_4 = 3$, iar excesul $E = 0$. Dacă excesul este pozitiv ($\alpha_4 > 3$; $E > 0$), curba este mai ascuțită decât curba normală, iar dacă este negativ ($\alpha_4 < 3$; $E < 0$), curba este mai turtită decât curba normală.

Dar, chiar la o abatere de la curba teoretică, se pot utiliza principiile teoretice specifice acesteia la determinarea parametrilor statistici ai curbei empirice. Însă, pentru aceasta, trebuie evaluate situațiile când este posibil acest lucru. De aceea au fost elaborate teste cu caracter teoretic care permit o evaluare corectă a condițiilor de utilizare a principiilor teoretice specifice pentru analiza statistică a datelor experimentale.

Verificarea normalității distribuției datelor experimentale. Există două procedee de bază pentru verificarea normalității distribuției datelor experimentale: utilizarea unui test de concordanță (χ^2 şi testul lui Kolmogorov); utilizarea reprezentării grafice a valorilor frecvențelor cumulate pe rețeaua de probabilitate de-a lungul unei drepte.

Se va considera testul χ^2. **Testul χ^2 se utilizează curent la verificarea concordanței între frecvențele empirice şi frecvențele teoretice în cazul celor mai diverse repartiții statistice.**

Funcția χ^2 a fost definită, de matematicianul englez Pearson Karl (1857-1936), sub forma:

$$\chi^2 = \sum \frac{(f_0 - f_c)^2}{f_c} \tag{3.81}$$

în care f_0 este frecvența observată, iar $\mathbf{f_c}$ este frecvența calculată. Observațiile se aranjează în grupe, numărul observațiile dintr-o grupă reprezentând frecvența grupei (intervalului). odată obținută suma (χ^2) pentru ansamblul intervalelor, se caută tabelar probabilitatea sa de apariție. În funcție de aceasta se decide. Se înțelege uşor că diferențele dintre $\mathbf{f_0}$ şi $\mathbf{f_c}$ trebuie să fie cât mai mici (la limită zero). Obişnuit χ^2_P se determină astfel încât:

Material cu drept de autor

$$P\left(\chi^2 > \chi_P^2\right) \leq 0{,}05 \qquad (3.82)$$

Date de observație se grupează pe intervale, numărul de date din fiecare interval trebuind să fie suficient de mare (cel puțin 5 sau şi mai bine 10). Pentru fiecare interval ($\mathbf{x_{i-1}}$-$\mathbf{x_i}$) se stabileşte numărul $\mathbf{n_i}$ de date ce cad în intervalul respectiv. După aceasta se caută probabilitatea p_i de a cădea în acest interval în ipoteza de normalitate a repartiției şi anume:

$$p_i = \Phi\left(\frac{x_i - \overline{x}}{S}\right) - \Phi\left(\frac{x_{i-1} - \overline{x}}{S}\right) = \Phi\left(z_i\right) - \Phi\left(z_{i-1}\right)$$

unde $\overline{x}$ este media aritmetică a măsurătorilor, **S** – abaterea medie pătratică (standard) de selecție; $\mathbf{z_i}$, $\mathbf{z_{i-1}}$ – variabilele normale normate corespunzătoare valorilor $\mathbf{x_i}$ şi $\mathbf{x_{i-1}}$. Probabilitatea $\Phi(z)$ şi $\Phi(z_i - 1)$ se caută în tabelul funcției Laplace.

Trebuie să se facă o precizare a modului cum se vor calcula probabilitățile $\mathbf{p_1}$ şi $\mathbf{p_k}$. Simbolul $\mathbf{p_1}$ reprezintă probabilitatea ca valorile caracteristicii să fie mai mici decât $\mathbf{x_1}$, limita superioară a primului interval. Prin urmare drept limită inferioară a acestui interval se consideră $-\infty$; cum însă $\Phi(-\infty) = -\Phi(+\infty) = -0{,}5$ rezultă că:

$$p_1 = \Phi(z_1) - (-0{,}5) = \Phi(z_1) + 0{,}5 \qquad (3.83)$$

De asemenea, $\mathbf{p_k}$ reprezintă probabilitatea ca valorile caracteristicii să aparțină intervalului **k**, dar pentru ca suma probabilităților să fie **1** trebuie să considere că ultimul interval are limita superioară $+\infty$. Or, $\Phi(+\infty) = 0{,}5$

$$p_k = 0{,}5 - \Phi\left(z_{k-1}\right) \qquad (3.84)$$

Expresia (3.81) devine:

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(n_i - n.p_i\right)^2}{n.p_i} \qquad (3.85)$$

în care **n** este numărul total al măsurătorilor, $\mathbf{n = \Sigma n_i}$, iar **k** numărul intervalelor. Valorile $\mathbf{n_i}$ sunt frecvențe absolute, iar probabilitățile $\mathbf{p_i}$ sunt frecvențe teoretice relative. Pentru a transforma în frecvențe teoretice absolute frecvențele calculate din expresiile (3.81) se înmulțesc cu numărul total al observațiilor **n**.

Exemplu de calcul 3.9. Într-un proces tehnologic de turnare a fontei maleabile se supun analizei chimice 130 şarje pentru a se constata dacă procentul de carbon se încadrează în limitele prestabilite.

După înregistrarea datelor şi divizarea pe intervale a câmpului de dispersie a procentului de carbon, s-au completat coloanele 1 şi 2 din tabelul 3.2. Datele astfel înregistrate se supun analizei statistice.

Material cu drept de autor

Printre altele, se cere a se stabili legea de distribuție a datelor experimentale. Care este aceasta ?

Rezolvare. Se efectuează reprezentarea grafică din figura 3.10.

Se emite ipoteza distribuției normale a datelor experimentale. În cazul distribuției normale, funcția de distribuție poate fi scrisă sub următoarea formă:

$$F(x) = 0{,}5 + \Phi(z) \qquad (3.86)$$

unde $\Phi(z)$ este funcția lui Laplace, iar $\mathbf{z = (x_i - \bar{x})/\sigma}$ - abaterea normată, $\mathbf{x_i}$ luând drept valori succesive limitele superioare ale intervalelor (1,79; 1,89; 1,99;...;2,79). Pentru a calcula pe **z** sunt necesare media şi abaterea medie pătratică al căror calcul se face cu ajutorul tabelului 3.1 şi tabelul 3.2.

Tabelul 3.1

Tabel de calcul pentru criteriul Kolmogorov

Limitele intervalelor	n_i	f_r	$F_n(x)$	$x - \bar{x}$	$z_i = \frac{x - \bar{x}}{\sigma}$	$\Phi(z)$	$F(x)=0{,}5+\Phi(z)$	$F_n(x)-F(x)$
1	2	3	4	5	6	7	8	9
1,70-1,79	1	0,0077	0,0077	-0,488	-2,53	-0,4943	0,0057	0,0020
1,80-1,89	6	0,0462	0,0539	-0,388	-2,01	-0,4778	0,0222	0,0317
1,90-1,99	6	0,0462	0,1001	-0,288	-1,49	-0,4319	0,0681	0,0320
2,00-2,09	8	0,0615	0,1616	-0,188	-0,97	-0,3340	0,1160	-0,0044
2,10-2,19	14	0,1077	0,2693	-0,088	-0,46	-0,1772	0,3228	-0,0535
2,20-2,29	34	0,2615	0,5308	-0,012	0,06	0,0239	0,5239	0,0069
2,30-2,39	28	0,2154	0,7462	-0,012	0,58	0,2190	0,7190	0,0272
2,40-2,49	19	0,1461	0,8923	0,212	1,10	0,3643	0,8643	0,0280
2,50-2,59	9	0,0692	0,9615	0,312	1,62	0,4474	0,9474	0,0141
2,60-2,69	4	0,0308	0,9923	0,412	2,13	0,4983	0,9983	-0,0060
2,70-2,79	1	0,0077	1,0000	0,512	2,65	0,5000	1,0000	
Total	130	1,0000						

$$\bar{x} = \frac{37}{130}0{,}1 + 2{,}25 = 2{,}278 \qquad \sigma = 0{,}1\sqrt{\frac{495}{130} - \left(\frac{37}{130}\right)^2} = 0{,}193$$

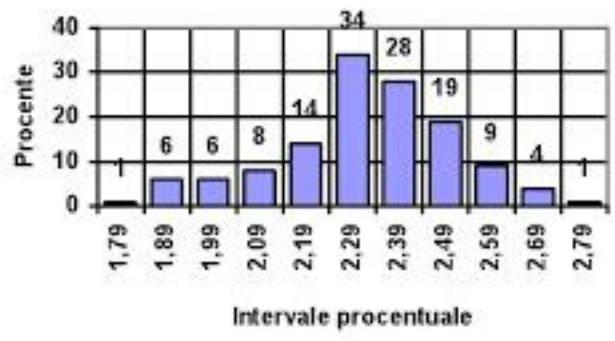


Fig.3.10. Histograma datelor statistice din tabelul 3.1

Pentru verificarea normalității distribuției datelor experimentale, adică a ipotezei emise, se va folosi criteriul lui Kolmogorov.

Aplicarea criteriului lui Kolmogorov conduce la stabilirea concordanței dintre distribuția teoretică şi distribuția empirică.

Astfel, se presupune că:

Material cu drept de autor

x_1	x_2	...	x_k
n_1	n_2	...	n_k

este distribuția empirică şi că se face ipoteza că variabila **X** are o distribuție a cărei funcție de frecvență este **f(x)**.

Se pot calcula frecvențele relative cumulate, adică valorile funcției de distribuție empirice:

$$F_n(x_1),\ F_n(x_2),\ldots,F_n(x_k) \tag{3.87}$$

unde:

$$F_n(x_i) = \frac{n_1 + n_2 + \ldots + n_i}{n} \tag{3.88}$$

precum şi valorile funcției de distribuție **F(x)** corespunzătoare valorilor x_i **(i = 1, 2,...k)**.

$$F(x_1),\ F(x_2),\ldots,F(x_3) \tag{3.89}$$

unde:

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x).dx \tag{3.90}$$

Tabelul 3.2

Conținut de carbon,%	Nr. şarjei n	x'	x'.n	$x'^2.n$
1,70-1,79	1	-5	-5	25
1,80-1,89	6	-4	-24	96
1,90-1,99	6	-3	-18	54
2,00-2,09	8	-2	-16	32
2,10-2,19	14	-1	-14	14
2,20-2,29	34	0	0	0
2,30-2,39	28	1	28	28
2,40-2,49	19	2	38	76
2,50-2,59	9	3	27	81
2,60-2,69	4	4	16	64
2,70-2,79	1	5	5	25
Total	130		37	495

Dacă caracteristica **X** urmează legea de distribuție **f(x)**, atunci frecvențele relative cumulate $F_n(x_i)$ au valori apropiate de valorile corespunzătoare ale funcției de distribuție $F(x_i)$, respectiv diferențele:

$$\left|F_n(x_1) - F(x_1)\right|;\left|F_n(x_2) - F(x_2)\right|;\ldots;\left|F_n(x_k) - F(x_k)\right| \tag{3.91}$$

sunt mici, adică nu vor depăşi o valoare determinată d_n. Această valoare trebuie determinată astfel încât probabilitatea **q** ca cea mai mare dintre diferențele (3.91) s-o depăşească (să fie mică, mai mică sau egală cu 0,05), adică:

Material cu drept de autor

$$P\left(d_n > \frac{\lambda}{\sqrt{n}}\right) = q \le 0{,}05 \qquad (3.92)$$

Dar,

$$P\left(d_n > \frac{\lambda}{\sqrt{n}}\right) = 1 - P\left(d_n < \frac{\lambda}{\sqrt{n}}\right) = 1 - K(\lambda) \qquad (3.93)$$

Unei probabilități date **q** îi corespunde prin relația (3.93) o valoare determinată λ_q aşa încât pentru o mărime a selecției **n** dată şi pentru un nivel de semnificație **q** dat se găseşte valoarea tabelară $d_n = \lambda_q / \sqrt{n}$.

Prin urmare, în cazul când distribuția empirică corespunde distribuției teoretice trebuie să fie satisfăcută relația:

$$P\left(d_n > \frac{\lambda_q}{\sqrt{n}}\right) = q \qquad (q \le 0{,}005) \qquad (3.94)$$

Dat fiind că probabilitatea evenimentului $d_n > \lambda_q / \sqrt{n}$ este mică, se poate considera că în condițiile corespondenței dintre distribuția empirică şi distribuția teoretică este imposibil ca cea mai mare diferență în valoare absolută dintre $\mathbf{F_n(x)}$ şi **F(x)** să fie mai mare decât mărimea $\lambda_q / \sqrt{n}$

Astfel că se pot formula următoarele reguli:

- dacă cea mai mare diferență dintre frecvențele relative cumulate şi valorile corespunzătoare ale funcției de distribuție teoretice este mai mică decât valoarea $\lambda_q / \sqrt{n}$, se conchide că variabila aleatoare **X** urmează legea de distribuție teoretică **f(x)**.

- dacă diferența $\mathbf{d_n = max|F_n(x)-F(x)|}$ este mai mare decât valoarea $\lambda_q / \sqrt{n}$, nu există nici un temei să se accepte ipoteza că variabila aleatoare **X** urmează legea de distribuție considerată **f(x)**.

Se va urmări aplicarea acestei metodici a criteriului lui Kolmogorov la exemplul considerat.

În coloana 6 din tabelul 3.1 sunt date valorile $\mathbf{z_i}$. În anexa 1 se caută valorile $\Phi(z)$ corespunzătoare. De remarcat că valorile $\Phi(z_i)$ corespunzătoare valorilor negative ale lui **z** sunt, de asemenea, negative. Aplicând formula (3.97) se calculează, în coloana 8 din tabelul 3.1, valorile funcției de distribuție $\mathbf{F(x_i)}$. În sfârşit, în coloana 9 sunt calculate diferențele $\mathbf{F_n(x_i)}$.

Se observă că cea mai mare diferență în valoare absolută este (-0,0535), deci $\mathbf{d_n = 0{,}0535}$. Această valoare se compară cu mărimea $\lambda_q / \sqrt{n}$.

Material cu drept de autor

Pentru **q = 0,025, K(λ) = 1 – q = 0,975** se găseşte $\lambda_{0,975}$ **= 1,48** aşa încât $\lambda_q / \sqrt{n} = 1,48/\sqrt{136} = 0,1298$.

Deoarece $\mathbf{d_n = 0,0535}$ este mai mic decât $\lambda_q / \sqrt{n} = 1,48/\sqrt{136} =$ 0,1298, se conchide că distribuția şarjelor de fontă maleabilă cenuşie, din punct de vedere al conținutului de carbon, urmează legea normală de distribuție.

Exemplu de calcul 3.10. În condițiile exemplului precedent, se cere a se verifica dacă conținutul de carbon al şarjelor de fontă maleabilă cenuşie urmează legea normală utilizându-se criteriul χ^2 **(HI – pătrat)**.

Rezolvare. Pentru a prezenta modul de aplicare a criteriului χ^2 **(HI – pătrat)**, se presupune, în general, că **n** valori observate se grupează în **k** grupe şi că $\mathbf{n_1, n_2,\ldots,n_k}$ reprezintă numărul observațiilor corespunzătoare acestor grupe, adică $\mathbf{n_1 + n_2 +\ldots+ n_k = n}$ şi fie $\mathbf{p_1, p_2,\ldots,p_k}$ probabilitățile ca fiecare din valorile observate să aparțină respectiv grupelor de mărime $\mathbf{n_1, n_2,\ldots,n_k}$ astfel încât $\mathbf{p_1 + p_2 +\ldots+ p_k = 1}$.

În cazul distribuției binomiale, aşa cum se cunoaşte, media şi dispersia acestei distribuții sunt **M[x] = n.p** şi **D[x] = n.p.q** aşa încât abaterea normată $z = \dfrac{x - n.p}{\sqrt{n.p.q}}$ se distribuie normal cu media **0** şi dispersia **1**, unde **x** reprezintă numărul de câte ori se realizează evenimentul a cărei probabilitate de realizare în orice încercare este egală cu **p**, iar **q** este probabilitatea de realizare a evenimentului contrar adică **q = 1 – p**. Urmează atunci că abaterile normate:

$$z_i = \frac{n_i - n.p_i}{\sqrt{n.p_i(1-p_i)}} \tag{3.95}$$

se distribuie normal cu media **0** şi dispersia **1** iar mărimea,

$$\chi^2 = \frac{(n_1 - n.p_1)^2}{n.p_1(1-p_1)} + \frac{(n_2 - n.p_2)}{n.p_2(1-p_2)} + \ldots + \frac{(n_k - n.p_k)^2}{n.p_k(1-p_k)} = \sum_{i=1}^{k} \frac{(n_i - n.p_i)}{n.p_i(1-p_i)} \tag{3.96}$$

are o distribuție χ^2 **(HI – pătrat)**.

Dacă între distribuția empirică şi distribuția teoretică există corespondență, atunci mărimea χ^2 **(HI – pătrat)** nu va depăşi o valoare dată χ_q^2 corespunzătoare nivelului de semnificație **q**, care trebuie astfel ales încât evenimentul $\chi^2 > \chi_q^2$ să fie practic imposibil. Practica arată că pentru

Material cu drept de autor

q trebuie aleasă o valoare mai mică sau cel mult egală cu **0,05** (**q** ≤ **0,05**). Deci χ_q^2 se determină astfel încât să fie satisfăcută relația:

$$P(\chi^2 > \chi_q^2) = q \leq 0{,}05 \qquad (3.97)$$

Se pot formula următoarele reguli pentru verificarea corespondenței dintre distribuția empirică şi distribuția teoretică. Dacă valoarea lui χ^2 determinată de relația:

$$\chi^2 = \frac{(n_1 - n.p_1)^2}{n.p_1} + \frac{(n_2 - n.p_2)}{n.p_2} + \ldots + \frac{(n_k - n.p_k)^2}{n.p_k} = \sum_{i=1}^{k} \frac{(n_i - n.p_i)}{n.p_i} \qquad (3.98)$$

cu **f = k – 1** grade de libertate este mai mică decât valoarea tabelată χ_q^2 corespunzătoare unui nivel de semnificație **q** ≤ **0,05**, se poate conchide că distribuția empirică urmează legea de distribuție **f(x)**.

Dacă valoarea χ^2 calculată pe baza datelor cercetării cu ajutorul formulei (3.98) este mai mare decât valoarea tabelară χ_q^2, nu există nici un temei să se considere că distribuția empirică urmează legea de distribuție **f(x)**.

S-a menționat că $\mathbf{p_i}$ reprezintă probabilitatea ca valorile observate să aparțină intervalului **i (i = 1, 2,…,k)**. Probabilitatea $\mathbf{p_i}$ se calculează în ipoteza că distribuția cercetată urmează o lege de distribuție bine determinată. De obicei însă distribuția ipotetică conține un număr de parametri care se estimează cu ajutorul distribuției care se compară cu distribuția teoretică. Aşa că nu se pot găsi valori exacte ale probabilităților $\mathbf{p_i}$, ci numai valori aproximative $\hat{p}_i$ care depind de datele cercetării. În acest caz, se utilizează drept criteriu mărimea:

$$\chi^2 = \sum \frac{(n_i - n.\hat{p}_i)}{n.\hat{p}_i} \qquad (3.99)$$

unde abaterile $(\mathbf{n_i} - \hat{p}_i.\mathbf{n_i})$ sunt legate printr-un număr mai mare de legături liniare. Acest număr depinde de numărul parametrilor care se estimează pe baza valorilor observate. Dacă numărul parametrilor estimați este egal cu ***l***, atunci numărul legăturilor liniare care se impun abaterilor $(\mathbf{n_i} - \hat{p}_i.\mathbf{n_i})$ este egal cu ***l*** **+ 1**. Prin urmare, mărimea determinată cu expresia (3.99) are o distribuție χ^2 cu **k – (*l* + 1)** grade de liberate. Deci, se poate spune că în aplicarea criteriului χ^2 la verificarea corespondenței dintre distribuția empirică şi distribuția teoretică se întâlnesc două situații:

Material cu drept de autor

1. forma distribuției este determinată, adică toți parametrii acesteia sunt cunoscuți şi;

2. parametrii distribuției teoretice se estimează pe baza datelor cercetării.

În primul caz, intervalul critic este definit de inegalitatea $\chi^2 > \chi_q^2$ unde χ^2 se determină cu ajutorul relației (3.96) şi are **k-1** grade de libertate. În cazul al doilea, intervalul critic este definit de inegalitatea $\chi^2 > \chi_q^2$ unde χ^2 se determină cu ajutorul formulei (3.99) şi are **k – (*l* + 1)** grade de liberate.

Şi într-un caz şi în celălalt probabilitățile $\mathbf{p_i}$ şi respectiv $\hat{p}_i$ se calculează urmând procedeul următor. Atât $\mathbf{p_i}$ cât şi $\hat{p}_i$ reprezintă probabilitatea ca valorile caracteristicii să aparțină grupei **i** sau să fie cuprinse în intervalul **i** care are limite valorile $\mathbf{x_{i-1}}$ şi $\mathbf{x_i}$. Dacă **F(x)** este funcția de distribuție, atunci $\mathbf{p_i}$ se obțin făcând diferența dintre valorile funcției de distribuție corespunzătoare valorilor $\mathbf{x_{i-1}}$ şi $\mathbf{x_i}$, adică:

$$p_i = F(x_i) - F(x_{i-1}) \qquad (3.100)$$

dacă parametrii din expresia funcției de distribuție sunt cunoscuți.

Dacă parametrii nu sunt cunoscuți, atunci ei se înlocuiesc cu estimațiile lor obținute pe baza datelor cercetării aşa încât în acest caz funcția de distribuție **F(x)** va fi înlocuită cu estimația $\hat{F}(x)$, iar probabilitățile $\hat{p}_i$ sunt:

$$\hat{p}_i = \hat{F}(x_i) - \hat{F}(x_{i-1}) \qquad (3.101)$$

Tehnica de calcule este aceeaşi atât pentru determinarea lui $\mathbf{p_i}$ cât şi a lui $\hat{p}_i$. Deosebirea constă în faptul că în primul caz parametrii au valori cunoscute, iar în al doilea caz se cunosc estimațiile lor, de aceea în cele ce urmează se va folosi relația (3.91). Trebuie însă să se aibă în vedere dacă în calcule se utilizează valori cunoscute ale parametrilor sau estimațiile lor deoarece acest lucru este important pentru stabilirea numărului gradelor de libertate ale mărimii χ^2.

În cele ce urmează se va arăta cum se utilizează criteriul χ^2 în cazul când se face ipoteza că variabila este distribuită normal.

În cazul când se face ipoteza că distribuția empirică urmează legea normală, funcția de distribuție are expresia:

Material cu drept de autor

$$F(x_i) = 0,5 + \Phi(z_i) \qquad (3.102)$$

unde $\Phi(z_i)$ este funcția tabelară a lui Laplace, $z_i = \frac{x_i - \bar{x}}{\sigma}$ în care x_i reprezintă limita superioară a intervalului **i**, iar $\bar{x}$ şi σ sunt parametrii distribuției, respectiv, media şi abaterea medie pătratică.

Rezultă că probabilitățile p_i se determină în acest caz cu ajutorul expresiei:

$$p_i = F(z_i) - F(z_{i-1}) = \Phi(z_i) - \Phi(z_{i-1})$$

Trebuie să se facă o precizare a modului cum se vor calcula probabilitățile p_1 şi p_k. Simbolul p_1 reprezintă probabilitatea ca valorile caracteristicii să fie mai mici decât x_1, limită superioară a primului interval. Prin urmare, drept limită inferioară a acestui interval se consideră $(-\infty)$ cum însă $\Phi(-\infty) = -\Phi(+\infty) = -0,5$, rezultă că:

$$p_1 = \Phi(z_1) - (-0,5) = \Phi(z_1) + 0,5$$

De asemenea, p_k reprezintă probabilitatea ca valorile caracteristicii să aparțină intervalului **k**, dar pentru ca suma probabilităților să fie **1** trebuie să se considere că ultimul interval are limita superioară $+\infty$. Or, $\Phi(+\infty) = 0,5$, aşa încât $p_k = 0,5 - \Phi(z_{k-1})$.

De remarcat că în cazul criteriului χ^2 trebuie să se facă o astfel de grupare încât frecvențele intervalelor să nu fie mai mici decât **5**.

Desfăşurarea calculelor se face conform schemei prezentate în tabelul **3.3**.

În tabelul **3.4** se prezintă calculul desfăşurat pentru χ^2, după schema de calcul din tabelul **3.3**.

Se observă că primele două şi ultimele două intervale au fost contopite pentru ca numărul valorilor aparținând noilor intervale să fie mai mare sau cel puțin egal cu **5**. Din tabel rezultă că $\chi^2 = 13,03$. Întrucât parametrii distribuției normale $\bar{x}$ şi σ s-au calculat pe baza datelor cercetării, χ^2 are $f = k - l - 1 = 9 - 2 - 1 = 6$ grade de libertate. Pentru nivelul de semnificație se alege valoarea $q = 0,025$. În anexa 2, corespunzător probabilității $P = 1 - q = 1 - 0,025 = 0,975$ şi numărul gradelor de libertate $f = 6$ se găseşte $\chi^2_{0,025} = 14,4$.

Valoarea $\chi^2 = 13,03$ calculată pe baza valorilor observate este mai mică decât valoarea tabelată $\chi^2_{0,025} = 14,4$, prin urmare se consideră distribuția conținutului de carbon al şarjelor de fontă maleabilă cenuşie urmează aproximativ legea de distribuție normală.

Material cu drept de autor

Tabelul 3.3

Tabel de calcul pentru aplicarea criteriului χ^2

Nr. crt	Limitele intervalului	n_i	$z_i = \frac{x_i - \bar{x}}{\sigma}$	$\Phi(z_i)$	$p_i = \Phi(z_i) - \Phi(z_{i-1})^*$	$n.p_i$	$(n_i - n.p_i)^2$	$\frac{(n_i - n.p_i)^2}{n.p_i}$
1	$x_0 - x_1$	n_1	z_1	$\Phi(z_1)$	p_1	$n.p_1$	$(n_1 - n.p_1)^2$	$\frac{(n_1 - n.p_1)^2}{n.p_1}$
2	$x_1 - x_2$	n_2	z_2	$\Phi(z_2)$	p_2	$n.p_2$	$(n_2 - n.p_2)^2$	$\frac{(n_2 - n.p_2)^2}{n.p_2}$
3	$x_2 - x_3$	n_3	z_3	$\Phi(z_3)$	p_3	$n.p_3$	$(n_3 - n.p_3)^2$	$\frac{(n_3 - n.p_3)^2}{n.p_3}$
...	...		...	...	...	...	...	...
i	$x_{i-1} - x_i$	n_i	z_i	$\Phi(z_i)$	p_i	$n.p_i$	$(n_i - n.p_i)^2$	$\frac{(n_i - n.p_i)^2}{n.p_i}$
...	...							
k	$x_{k-1} - x_k$	n_k	z_k	$\Phi(z_k)$	p_k	$n.p_k$	$(n_k - n.p_k)^2$	$\frac{(n_k - n.p_k)^2}{n.p_k}$
Total		n			1,00			$\chi^2 = \frac{(n_i - n.p_i)^2}{n.p_i}$

Material cu drept de autor

Tabelul 3.4

Tabel de calcul pentru aplicarea criteriului χ^2

Limitele intervalului	n_i	$x_i - \bar{x}$	$z_i = \frac{x_i - \bar{x}}{\sigma}$	$\Phi(z_i)$	$p_i = \Phi(z_i) - \Phi(z_{i-1})^*$	$n.p_i$	$(n_i - n.p_i)$	$(n_i - n.p_i)^2$	$\frac{(n_i - n.p_i)^2}{n.p_i}$
1,70 – 1,79									
1,80 – 1,89	7	-0,388	2,01	-0,4778	0,0222	2,89	4,11	16,8921	5,85
1,90 – 1,99	6	-0,288	-1,49	-0,4319	0,0459	5,97	0,03	0,0009	0,00
2,00 – 2,09	8	-0,188	-0,97	-0,3340	0,0979	12,73	-4,73	22,3729	1,76
2,10 – 2,19	14	-0,088	-0,46	-0,1772	0,1568	20,38	-6,38	40,7044	2,00
2,20 – 2,29	34	0,012	0,06	0,0239	0,2011	26,14	7,86	61,7796	2,36
2,30 – 2,39	28	0,112	0,58	0,2190	0,1951	25,36	2,64	6,9696	0,27
2,40 – 2,49	19	0,212	1,10	0,3643	0,1453	18,89	0,11	0,0121	0,00
2,50 – 2,59	9	0,312	1,62	0,4474	0,0831	10,80	-1,80	3,2400	0,30
2,60 – 2,69									
2,70 – 2,79	5	0,512	2,65	0,5000	0,0526	6,84	-1,84	3,3856	0,49
	130								13,03

Material cu drept de autor

Cap.4. POPULAȚIE STATISTICĂ ŞI EŞANTION STATISTIC

4.1. Cercetări selective: de la populație la eşantion

Obiectivul legitim al cercetării ştiințifice este identificarea unor adevăruri cu un anumit grad de generalitate. Din punct de vedere statistic „generalul" este reprezentat de totalitatea valorilor care descriu o anumită caracteristică, şi este numit „populație". Din păcate însă, investigarea tuturor „indivizilor" (valorilor) care compun o anumită populație nu este aproape niciodată posibilă. Ca urmare, în practica cercetării ştiințifice se supun cercetării loturi mai restrânse, extrase din ansamblul colectivității vizate, ai căror parametri descriptivi (medie, variabilitate) sunt extrapolați, în anumite condiții şi cu ajutorul unor proceduri specializate, la populația din care fac parte.

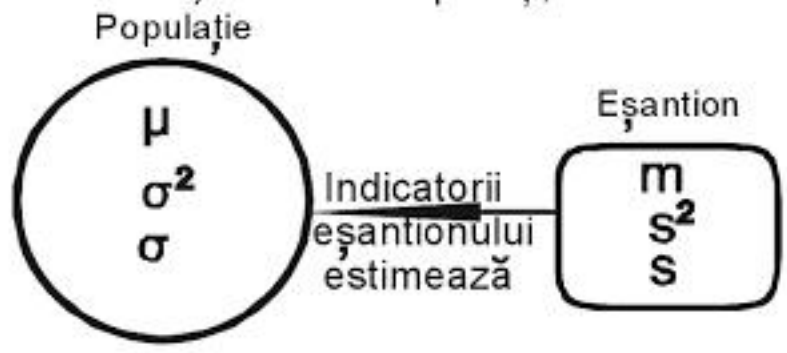


Fig.3.1. Parametrii statistici definiți pentru populație şi eşantion

A fundamenta un adevăr statistic înseamnă a trage o concluzie care descrie parametrii unei populații de valori, pe baza indicatorilor unui eşantion din acea populație.

În contextul cercetării statistice utilizăm următoarele definiții:

Populație - **totalitatea unităților de informație care constituie obiectivul de interes al unei investigații**. Prin „unități individuale de informație" înțelegem elementele individuale constitutive ale populației.

Eşantionul, reprezintă **unitățile de informație selecționate pentru a fi efectiv studiate**. Ideea pe care se bazează cercetările pe eşantioane, este aceea că se pot face aprecieri asupra unei întregi populații, în anumite condiții, doar pe baza caracteristicilor măsurate pe o parte a acesteia.

4.2. Reprezentativitatea eşantionului

Decizia de a culege datele necesare unei cercetări de la un eşantion sau de la o populație depinde de o serie de aspecte practice. Astfel, în unele situații, dacă timpul, resursele financiare şi umane nu constituie o problemă sau dacă populația țintă nu este foarte numeroasă, atunci este mult mai avantajoasă culegerea datelor de la toți indivizii care compun o populație vizată; în felul acesta se obține o imagine exactă a problematicii investigate. În alte situații există o serie de constrângeri care îl împiedică pe cercetător să ajungă la toți indivizii care compun o populație.

Setul de operații cu ajutorul cărora, din ansamblul *populației* vizate de cercetare, se extrage o parte, numită *eşantion*, parte ce va fi

Material cu drept de autor

supusă nemijlocit investigației este desemnat ca fiind operația de **eşantionare**.

Calitatea unui eşantion de a permite extinderea concluziilor la întreaga populație din care a fost extras se numeşte reprezentativitate.

Modul de constituire a eşantionului este decisiv pentru nivelul de reprezentativitate. Esențială în acest caz este asigurarea condițiilor ca acesta să acopere în mod real caracteristicile populației, evitându-se „favorizarea" sistematică a unor subiecți „nereprezentativi". Fără a intra în amănunte tehnice cu privire la procedurile de eşantionare iată care sunt, principial, cele mai utilizate metode de constituire a eşantioanelor:

a. **Eşantionare stratificată multistadială**. Populația se împarte în categorii, fiecare categorie în subcategorii ş.a.m.d., iar subiecții sunt selecționați aleator la nivelul categoriei de nivelul cel mai scăzut. Se obține astfel un eşantion care reproduce fidel structura populației.

b. **Eşantionare prin clasificare unistadială**. Se identifică categorii pe un singur nivel iar subiecții se extrag aleator din fiecare categorie.

c. **Eşantionare aleatoare**. Subiecții sunt extraşi aleator (la întâmplare) din ansamblul populației. „La întâmplare", înseamnă în acest caz utilizarea unei proceduri care asigură fiecărui subiect al populației absolut aceleaşi şanse de a fi inclus în eşantion. În acest scop se pot utiliza programe de calculator sau tabele de numere aleatoare.

d. **Eşantionare pseudo-aleatoare (de convenienţă)**. Sunt utilizați subiecții „disponibili". Este cazul cel mai frecvent întâlnit în practică şi, dacă „disponibilitatea" nu este afectată de un aspect care să influențeze semnificativ obiectivul cercetării, atunci reprezentativitatea este acceptabilă.

4.3. Erorile cercetării statistice prin sondaj

În accepțiunea cea mai largă, se consideră eroare de selecție abaterea care există între valoarea unui parametru (de exemplu, media) calculat prin prelucrarea datelor din eşantion şi valoarea aceluiaşi parametru care s-ar fi obținut dacă s-ar fi organizat o observare totală şi ar fi fost prelucrate datele de la toate unitățile colectivității.

Erorile întâlnite în cadrul sondajului sunt de două feluri:

- erori comune tuturor tipurilor de observări - ***erori de înregistrare;***
- erori specifice cercetării prin sondaj - ***erori de reprezentativitate***.

Deoarece înregistrarea datelor se face de un personal specializat şi pentru un număr restrâns de unități, de regulă, în sondaje, erorile de înregistrare apar într-un număr mic de cazuri şi pot fi înlăturate cu uşurință printr-un eventual control riguros.

Erorile de reprezentativitate specifice sondajului pot fi de două feluri: *erori sistematice şi erori întâmplătoare*.

Material cu drept de autor

Erorile de reprezentativitate sistematice pot fi evitate dacă se respectă întocmai principiile teoriei selecției, prin înlăturarea cauzelor ce duc la producerea lor. Principalele cauze care pot duce la apariția erorilor sistematice sunt:

- alegerea deliberatã a aşa-ziselor unități "reprezentative";
- alegerea la "nimerealã" (nu la întâmplare) a unitãþilor de eşantion;
- selectarea preferențială a acelor unități care să ducă la rezultatul dorit de cercetător;
- substituirea din comoditate a unei unități de cercetare prin alta asemănătoare;
- cuprinderea incompletă în sondaj a unităților, din motive de comoditate.

Erorile întâmplătoare de reprezentativitate pot apare chiar dacă se respectă cu strictețe aceste reguli. Ele derivă din însăşi esența metodei de cercetare prin sondaj. Prin numărul mic de unități care alcătuiesc eşantionul nu se poate reproduce decât întâmplător identic seria de distribuție a variabilei din colectivitatea generalã sau parametrii acesteia.

Deşi nu pot fi evitate, erorile de reprezentativitate, pot fi calculate cu anticipație dacă selecția este probabilistică. Estimarea parametrilor din colectivitatea generală se va putea face deci pe baza indicatorilor obținuți din prelucrarea datelor de sondaj cu o eroare întâmplătoare de reprezentativitate care se găseşte într-un anumit interval probabilistic.

Rezultă deci că fiecărui indicator derivat sau sintetic trebuie să i se ataşeze şi eroarea sa de reprezentativitate, pentru a putea fi generalizat pentru întregul ansamblu.

În practica sondajului erorile de reprezentativitate se pot calcula ca ***erori efective*** şi ca ***erori probabile.***

4.3.1 Erori efective. Verificarea reprezentativității eşantionului

Erorile efective de reprezentativitate se pot calcula numai pentru caracteristicile la care s-au obținut date şi dintr-o observare totală.

Considerând că şi în acest caz media este indicatorul sintetic cel mai reprezentativ, eroarea efectivă de sondaj se calculează ca diferență între media eşantionului şi media colectivității totale. Calculul erorii efective de sondaj nu este altceva decât verificarea gradului de reprezentativitate a unui eşantion în raport cu structura colectivității totale.

Practica demonstrează că numai întâmplător este posibilă reproducerea unei structuri identice cu aceea a colectivității generale, ceea ce înseamnă că eşantioanele extrase pot avea grade diferite de reprezentativitate. Ca atare, înainte de a se trece la prelucrarea datelor culese se va face verificarea reprezentativității eşantionului.

Caracteristicile utilizate la alegerea eşantionului prezintă, de regulă, forme variate de manifestare şi de aceea verificarea reprezentativității eşantionului nu este întotdeauna o operație uşoară.

Material cu drept de autor

La verificarea reprezentativității eşantionului se porneşte de la compararea structurii pe grupe a colectivității de selecție cu cea a colectivității generale, denumită şi structură programată. În cazul în care această structură nu diferă cu mai mult de +/- 5% se acceptă eşantionul constituit ca fiind reprezentativ.

4.3.2 Eroarea medie probabilă şi eroarea limită

Întreaga teorie a sondajului statistic se bazează pe principiile teoriei probabilităților şi ale statisticii matematice, de la formarea eşantionului şi până la estimarea parametrilor colectivității totale pe baza datelor din eşantion.

În teoria selecției se demonstrează că dacă volumul colectivității de selecție este suficient de mare, mediile de sondaj urmează la limită legea distribuției din colectivitatea generală, iar media de selecție ca expresie sintetică a nivelurilor individuale ale tuturor unităților cercetate, va fi cât mai aproape de media colectivității generale.

În practică însă pentru acelaşi volum de selecție se pot obține mai multe eşantioane extrase succesiv din aceeaşi colectivitate totală, obținând astfel valori diferite ale mediei de selecție. În acest proces de formare a mediilor de selecție fiecare medie poate să apară o singură dată sau de mai multe ori. Se confirmă astfel că şi media de selecție este tot o variabilă aleatoare căreia i se poate stabili legea de distribuție.

Pentru a putea urmări modul de formare a distribuției de eşantionare a mediei se va lua un exemplu, în care volumul colectivității generale este de patru unități (**A; B; C; D**) din care se vor forma toate eşantioanele posibile din câte două unități, deci **N=4**, iar **n=2**.

Dacă se efectuează toate selecțiile posibile, folosind procedeul bilei revenite (selecție repetată) se vor obține următoarele eşantioane, de câte două unități:

(A+A); (A+B); (A+C); (A+D); (B+A); (B+B); (B+C); (B+D);
(C+A); (C+B); (C+C); (C+D); (D+A); (D+B); (D+C); (D+D).

Generalizând, înseamnă că în cazul selecției repetate se poate efectua un număr de eşantioane egal cu N^n, respectiv în exemplul luat $4^2=16$.

În cazul selecției nerepetate (procedeul bilei nerevenite) combinațiile sunt mai puține datorită faptului că aceeaşi unitate nu poate participa decât într-un singur eşantion.

În exemplul prezentat vor apare următoarele combinații posibile:

(A+B); (A+C); (A+D); (B+A); (B+C); (B+D); (C+D). Formula de calcul a numărului de eşantioane posibile pentru selecția nerepetată este dată de formula de calcul a combinărilor, aplicată la selecție:

$$C_N^n = \frac{N!}{n!(N-n)!} \qquad (4.1)$$

Material cu drept de autor

Fiecare eşantion va fi definit de o medie şi o dispersie calculabilă pentru fiecare caracteristică înregistrată, care vor prezenta abateri față de media şi dispersia colectivității totale.

Erorile de selecție obținute ca diferențe între media de selecție şi media generală iau valori diferite de la un eşantion la altul, ceea ce face necesară calcularea unui indicator sintetic numit ***eroarea medie de reprezentativitate.***

Pentru a evita compensarea unor erori de sensuri diferite la calculul erorii medii de reprezentativitate se foloseşte media pătratică a abaterilor, notată cu $\sigma_{\bar{x}}$:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum_{s=1}^{k}(\bar{x}_s - \bar{x}_0)^2 . n_s}{\sum_{s=1}^{k} n_s}} \qquad (4.2)$$

în care : **k** - reprezintă numărul eşantioanelor posibile; $\mathbf{n_s}$ - frecvența mediilor de selecție posibile.

Dacă se consideră media de selecție în sens probabilistic, adică fiecare valoare a sa ca eveniment favorabil față de toate celelalte evenimente egal posibile, oricât am repeta selecția, frecvența de apariție a fiecărei valori a mediei, luată ca frecvență relativă, are tendința de a coincide cu probabilitatea de producere a erorii sale față de valoarea mediei colectivității generale. În cazul selecției aleatoare se poate demonstra că pentru o anumită fracție de selecție (**n/N**), fiecărei valori a mediei de selecție îi corespunde o probabilitate de apariție bine determinată, în funcție de mărimea absolută a abaterii sale față de media colectivității generale. De aici rezultă că mărimea şi probabilitățile de apariție a diferitelor medii de selecție sunt determinate şi de volumul eşantionului. Potrivit legii numerelor mari, cu cât eşantionul cuprinde un număr mai mare de unități în raport cu cel al colectivității generale cu atât media de selecție va estima mai bine media pe total.

Aceasta corespunde legii numerelor mari, formulată de J.Bernoulli, potrivit căreia probabilitatea ca diferența în valoare absolută dintre frecvența relativă **f*** şi probabilitatea **p** de producere a unui eveniment să fie mai mică decât un număr pozitiv şi arbitrar **ε**, tinde către unu când volumul eşantionului tinde către infinit, adică:

$$\lim_{n\to\infty} P(f^* - p| < \varepsilon) \to 1 \quad oricare \quad ar\ fi\ \varepsilon > 0$$

Respectând cerințele legii numerelor mari, în teoria selecției se demonstrează că dacă volumul eşantionului este suficient de mare (pentru fenomenele monotipice peste 40 de unități), media de selecție se distribuie

Material cu drept de autor

potrivit funcției Gauss - Laplace cunoscută în statistică sub denumirea de distribuție normală.

Distribuția normală este de forma unei distribuții simetrice în care cea mai mare probabilitate de apariție în cazul sondajului o are acea medie de selecție care coincide în valoare cu media colectivității generale şi pentru care eroarea de reprezentativitate este egală cu zero. Față de această valoare centrală, celelalte valori ale mediei de selecție se distribuie simetric de ambele părți cu probabilități bine determinate, egale pentru aceeaşi abatere absolută într-un sens sau altul. Față de probabilitatea maximă, probabilitățile de apariție a mediilor de selecție descresc proporțional şi simetric, către capetele distribuției. Deci se poate spune că probabilitățile descresc pe măsură ce cresc erorile de reprezentativitate care sunt exprimate în aceleaşi unități de măsură ca şi variabila studiată. Pentru a le elibera de această formă concretă şi a le face comparabile pentru orice variabilă numerică, abaterile absolute se transformă în abateri normale normate:

$$\frac{\bar{x}_s - \bar{x}_0}{\sigma_{\bar{x}}} = z_i \qquad (4.3)$$

Dacă mediile de selecție se distribuie după legea normală înseamnă că şi erorile întâmplătoare de reprezentativitate urmează aceeaşi formă de repartiție şi interpretarea lor se face pe baza proprietăților distribuției normale (vezi tabelul 4.1), potrivit căreia trebuie stabilit intervalul de încredere, nivelul de siguranță şi pragul de semnificație.

Tabelul 4.1

Intervalele de încredere, nivelurile de siguranță şi pragurile de semnificație

Intervalul de încredere $(\bar{x}_s \pm z.\sigma_{\bar{x}})$	Nivelul de siguranță (%)	Pragul de semnificație (%)
$(\bar{x}_s \pm \sigma_{\bar{x}})$	68,26	31,74
$(\bar{x}_s \pm 1{,}96.\sigma_{\bar{x}})$	95,00	5,00
$(\bar{x}_s \pm 2.\sigma_{\bar{x}})$	95,44	4,56
$(\bar{x}_s \pm 2{,}58.\sigma_{\bar{x}})$	99,00	1,00
$(\bar{x}_s \pm 3.\sigma_{\bar{x}})$	99,73	0,27
$(\bar{x}_s \pm 4.\sigma_{\bar{x}})$	99,99	0,00

Material cu drept de autor

Pentru interpretarea modului de formare a erorii de reprezentativitate, pentru o selecție probabilistă de **n** unități extrase dintr-o colectivitate generală **N**, a cărui volum este suficient de mare şi pentru care se poate formula ipoteza că media de selecție se distribuie normal se foloseşte şi reprezentarea grafică, figura 4.1.

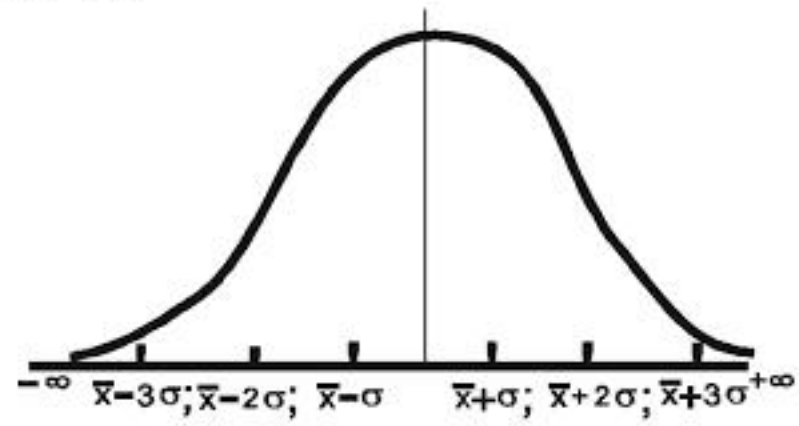


Fig.4.1. Distribuția normală

Interpretând graficul se constată că pe diferite intervale de variație a mediei de selecție, considerate ca intervale de încredere, se operează cu două noțiuni complementare: nivelul de siguranță şi pragul de semnificație.

Dacă se efectuează toate selecțiile posibile, intervalul de variație al tuturor mediilor obținute este acoperit cu suma probabilităților lor de apariție $\sum_{s=1}^{k} p_s = 1{,}0$; unde k este numărul tuturor mediilor de selecție. Dacă se restrânge intervalul de variație a mediei de selecție, atunci eroarea de reprezenntavitate se poate calcula în abateri normale normate ale mediei de selecție de la media colectivității totale. Aceste intervale de variație $(\bar{x}_0 \pm \sigma_{\bar{x}})$; $(\bar{x}_0 \pm 2.\sigma_{\bar{x}})$; $(\bar{x}_0 \pm 3.\sigma_{\bar{x}})$... sunt garantate ca nivel de siguranță cu probabilitățile corespunzătoare funcției de repartiție normale.

Potrivit teoremei formulată de Leapunov, probabilitatea ca variabila aleatoare (media de eşantionare) să fie cuprinsă între două limite fixate anticipat, adică: $P = (\bar{x}_s - z_\alpha.\sigma_{\bar{x}}) < \bar{x}_0 < (\bar{x}_s + z_\alpha.\sigma_{\bar{x}})$ poate fi aproximată cu relația:

$$\frac{1}{\sqrt{2.\pi}} \int_{-z_\alpha}^{z_\alpha} e^{-\frac{z^2}{2}}.dz = \Phi(z_\alpha)$$

Produsul $z_\alpha.\sigma_{\bar{x}}$ este cunoscut sub denumirea de eroare limită (Δ_x):

$$\Delta_{\bar{x}} = z_\alpha.\sigma_{\bar{x}}$$

Coeficientul $\mathbf{z_\alpha}$ reprezintă argumentul funcției Laplace şi se găseşte tabelat. Urmărind valorile probabilității **Φ(z)**, se constată că $\mathbf{z_\alpha}$ creşte corespunzător (vezi tabela anexă), ceea ce înseamnă că, cu cât scade probabilitatea cu atât creşte intervalul de încredere al mediei şi scade exactitatea cu care se estimează media generalã pe baza mediei de selecție.

Material cu drept de autor

Calculul erorii medii de reprezentativitate, necesită cunoaşterea tuturor mediilor de selecție posibile, frecvențelor corespunzătoare lor şi a mediei colectivității generale.

În practică însă cercetarea prin sondaj se foloseşte fie pentru a completa o observare totală de mare amploare, fie ca singura posibilitate de caracterizare statistică a fenomenelor respective. În aceste condiții, cunoscând, de regulă, numai media calculată pe baza datelor rezultate de la un singur eşantion, pentru determinarea erorii medii de reprezentativitate trebuie să se recurgă la relația care există între abaterea medie pătratică (σ_0), eroarea medie pătratică de reprezentativitate ($\sigma_{\bar{x}}$) şi volumul eşantionului.

În teoria selecției se demonstrează că în cazul ***selecției aleatoare repetate*** între cei doi indicatori amintiți există relația: $\sigma_0^2 = \sigma_{\bar{x}}^2 . n$ de unde:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma_0^2}{n}} = \frac{\sigma_0}{\sqrt{n}} \qquad (4.4)$$

Se desprinde concluzia că eroarea medie de reprezentativitate este direct proporțională cu abaterea medie pătratică a colectivității generale şi invers proporțională cu radical din volumul eşantionului.

Cum abaterea medie pătratică, în anumite condiții de timp şi spațiu este un indicator stabil, rezultă că mărimea erorii medii de selecție poate fi influențată în plus sau în minus prin modificarea volumului eşantionului. De regulă, urmărindu-se reducerea erorii de reprezentativitate se foloseşte relația:

$$\frac{\sigma_{\bar{x}}}{k} = \sqrt{\frac{\sigma_0^2}{k^2 . n}} \qquad (4.5)$$

În cazul selecției nerepetate numărul de eşantioane fiind mai mic datoritã faptului că fiecare unitate nu participă decât o singură dată la selecție, relația $\sigma_0^2 = \sigma_{\bar{x}}^2 . n$ se transformă într-o inegalitate deoarece câmpul de variație al erorilor de reprezentativitate se micşorează.

În relațiile existente între indicatorii variației din colectivitatea generală şi cei de selecție, în acest caz va interveni raportul **(N-n)/(N-1)**. Cu acest raport se corectează eroarea medie pentru sondajul repetat şi se obține formula de calcul a erorii medii pentru sondajul nerepetat:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma_0^2}{n}\left(\frac{N-n}{N-1}\right)} \qquad (4.6)$$

Material cu drept de autor

În practică, dacă volumul colectivității generale este foarte mare se renunță la (**-1**) din numitorul formulei şi formula de calcul a erorii medii de reprezentativitate devine:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma_0^2}{n}\left(1-\frac{n}{N}\right)} \qquad (4.7)$$

Pentru caracteristica alternativă, eroarea medie de reprezentativitate se va nota σ_w, deci:

- pentru selecția repetată:

$$\sigma_w = \sqrt{\frac{p.(1-p)}{n}} \qquad (4.8)$$

- pentru selecția nerepetată:

$$\sigma_w = \sqrt{\frac{p.(1-p)}{n}}.\sqrt{\left(1-\frac{n}{N}\right)} \qquad (4.9)$$

Dacă nu se dispune de dispersia din colectivitatea generală, atunci se poate înlocui cu un estimator al acesteia, adică dispersia de eşantion. În acest caz se corectează volumul eşantionului din formula erorii medii cu un grad de libertate.

Calculul erorii medii de reprezentativitate presupune efectuarea unor calcule necesare determinării erorii limită, ce se poate accepta în cercetarea prin sondaj.

Deoarece media sondajului folosit pentru estimarea parametrilor din colectivitatea generală poate lua valori mai apropiate sau mai îndepărtate de media generală, este necesar să se stabilească mărimea erorii limită. Eroarea limită de reprezentativitate se determină ca o abatere între media de selecție şi media colectivității generale, garantată cu probabilitatea corespunzătoare limitelor intervalului de variație.

Din formulele prezentate se observă că eroarea limită este o mărime variabilă, direct proporțională cu probabilitatea cu care se garantează rezultatele şi invers proporțională cu precizia rezultatelor.

Eroarea limită se poate mări fie prin modificarea volumului eşantionului (**n**), fie prin modificarea probabilității cu care se garantează rezultatele, deoarece pentru condiții date de timp şi spațiu dispersia colectivității totale este o valoare constantă.

În concluzie, **se poate afirma că eroarea medie şi eroarea limită pot fi calculate *anticipat*** dacă despre o variabilă statistică se cunosc media şi dispersia generală sau un estimator al acestora şi s-a stabilit cu ce volum al eşantionului se va lucra şi cu ce probabilități se vor garanta rezultatele.

Pentru acelaşi volum al eşantionului se obține o singură eroare medie şi mai multe erori limită schimbând coeficientul de probabilitate şi

Material cu drept de autor

pentru acelaşi coeficient de probabilitate se pot obține mai multe erori limită dacă se modifică volumul eşantionului. Deci, valoarea erorii limită depinde de volumul de selecție şi de siguranța cu care se estimează abaterea dintre media eşantionului şi media colectivității generale.

Indicatorii de selecție calculați – media, eroarea medie de reprezentativitate şi eroarea limită – pot servi la estimarea parametrilor din colectivitatea generală. În acest scop se folosesc următoarele procedee:

- procedeul coeficientului de corectare a erorilor de înregistrare;
- procedeul extinderii directe a rezultatelor sondajului la estimarea colectivității totale.

Procedeul coeficientului de corectare a erorilor de înregistrare se bazează pe probabilitatea ca erorile depistate în sondajul efectuat să coincidă cu aceeaşi probabilitate ca şi pe total. Deci, refăcând înregistrarea la o parte a unităților selectate aleator se face raportul dintre datele observării totale şi cele de sondaj şi coeficientul respectiv se aplică datelor din observarea totală.

Procedeul extinderii directe este cel mai frecvent utilizat în aplicarea cercetării prin sondaj ca mijloc de caracterizare a întregii colectivități. Aplicarea acestui procedeu permite estimarea intervalului de încredere pentru nivelul mediu al caracteristicii cât şi determinarea limitelor între care variază nivelul totalizat al caracteristicii pe întreaga colectivitate, acest lucru fiind posibil numai în cazul în care se cunoaşte volumul colectivității totale (**N**).

De asemenea, pornind de la relația de calcul a erorii limită se poate determina volumul eşantionului. Formulele de calcul diferă în funcție de tipul selecției şi de procedeul de selecție folosit.

4.4. Determinarea mărimii eşantionului

Determinarea volumului eşantionului este pasul esențial ce trebuie parcurs înainte de culegerea datelor. Volumul eşantionului este desemnat prin numărul unităților statistice simple sau complexe ce vor fi prelevate din populația de referință, de la care se vor înregistra datele de intrare pentru analiză.

Deci, analistul trebuie să decidă, în funcție de mai mulți factori, care este numărul optim de unități statistice ce trebuie cuprinse în sondaj pentru ca eşantionul să fie reprezentativ şi rezultatele sale să se poată extinde asupra populației de referință cu respectarea principiilor inferenței statistice.

În consecință, se pune problema determinării dimensiunii optime a eşantionului care să asigure îndeplinirea obiectivelor sondajului. Eşantionul ideal trebuie să fie în concordanță cu eşantionul practic, ce poate fi construit astfel încât să se poată atinge obiectivele studiului.

Un alt element care trebuie foarte clar precizat şi utilizat este reprezentat de modalitatea de determinare a volumului eşantionului. Contrar

Material cu drept de autor

simțului comun, nu este importantă ponderea eşantionului în colectivitatea totală (care intră doar ca element de corecție), cât mărimea absolută a lui.

Fixarea aprioricâ a proporției de sondaj, raportul dintre volumul eşantionului şi volumul populației totale, poate conduce fie la supra-dimensionare, fie la sub-dimensionare.

O rație de sondaj de zeci de procente se dovedeşte nesemnificativă în condițiile unui volum relativ redus al colectivității totale, după cum o pondere chiar mai mică de 1% este suficientă la colectivități mari şi foarte mari.

Volumul optim de eşantionare depinde de obținerea volumului minim care să asigure o reprezentativitate adecvată a eşantionului, ce este dat de factori de influență situați în afara mărimii colectivității totale şi care se referă la structura colectivității.

Expresiile de definiție a mărimii eşantionului sunt:

$$n = t^2 . \sigma^2 / \Delta_x^2 \qquad (4.10)$$

pentru caracteristici continue, şi, respectiv:

$$n = t^2 . P . (100 - P) / \Delta_x^2 \qquad (4.11)$$

pentru caracteristici alternative,

unde: **t** - valoarea teoretică corespunzătoare probabilității cu care se lucrează (de regulă, **P** = 95%, iar t = 1,96); **σ** - abaterea medie pătratică a distribuției caracteristicii care stă la baza elaborării eşantionului (σ^2 = dispersia sau varianța **V**); **P** - procentul în care populația cercetată posedă caracteristica de eşantionare; Δ_x - eroarea limită de reprezentativitate admisă.

Datorită faptului că nu se studiază întreaga colectivitate, estimarea valorilor obținute (medii, procente) la nivelul eşantionului pentru întreaga colectivitate se face cu o anumită eroare. Valoarea reală se află cuprinsă în limitele determinate de mărimea obținută la nivelul eşantionului ± Δ_x.

În determinarea volumului eşantionului se acceptă aprioric o anumită eroare cuprinsă între 1% (foarte rar practicată, întrucât necesită eşantioane foarte mari) şi 5% (prag de eroare aproape general acceptat de către cercetători).

În continuare, algoritmul de lucru este „simplu". La o anumită valoare a lui **σ** sau **P** şi o valoare impusă a lui Δ_x rezultă în mod automat o anumită valoare a lui **n**. Dificultatea cea mai mare constă însă tocmai în obținerea informațiilor referitoare la distribuția caracteristicilor de eşantionare, respectiv valorile ***σ*** sau **P**. La valorile **σ = 0,5 m** sau **P = 50%** şi Δ_x = ***1%*** rezultă un eşantion de 9600 unități statistice necesar pentru a fi studiat, ori, de multe ori, colectivitatea totală este mică.

Volumul cerut pentru eşantion scade vertiginos dacă ne reducem dorința de rigurozitate şi acceptăm valori mai mari ale lui Δ_x, ceea ce se şi face de regulă. Pentru un nivel al erorii Δ_x = 2%, 3%, 4% sau 5% (menținând

Material cu drept de autor

constant **σ** sau **P**) volumul eşantionului se reduce şi el la: 2400, 1060, 600 şi, respectiv, 384 unităţi statistice. Dacă se modifică valorile **σ** sau **P** vor rezulta mărimi diferite pentru **n**.

În evaluarea gradului de reprezentativitate a cercetărilor intervin şi alte elemente, care se referă la aplicarea corectă a tuturor cerinţelor de întocmire a eşantionului, a corecţiilor cerute de schema de eşantionare, de dispersia spaţială a populaţiei, de selectarea subiecţilor şi de cercetarea efectivă a acestora.

Putem preciza o serie de principii ce trebuie respectate pentru evaluare: aplicarea eşantionării la colectivităţile de populaţie reclamă asigurarea unui număr minim de persoane (**n**) care să permită un grad acceptabil de reprezentativitate; mărimea colectivităţii totale intervine doar ca element de corecţie, factorul de corecţie fiind **(N – n)/(N – 1)**, cu care se micşorează valoarea **n** a volumului eşantionului, deja obţinută. Valoarea raportului se apropie de unu în cazul în care **N** este un număr mare.

Atunci când studiem colectivităţi relativ mici (sub 500 unităţi statistice) nu se pot determina eşantioane reprezentative după regulile numerelor mari şi ar trebui să se ia în considerare cerinţele suplimentare ale eşantioanelor mici. Când suntem în imposibilitatea de a construi eşantioane reprezentative este preferabil să studiem loturi omogene sau neomogene de populaţie, dar care nu ne permit să generalizăm rezultatele la ansamblul colectivităţii, de unde rezultă implicit avantajele cercetării selective, riguros realizate.

O atenţie cu totul deosebită trebuie acordată modului în care se face uz de o metodă mai rapidă de determinare a mărimii eşantionului, care pleacă de la volumul colectivităţii totale (**N**) fără a mai lua în considerare caracteristicile populaţiei, *expresia Taro Jamane*:

$$N = n/(1 + N.\Delta_x^2) \qquad (4.12)$$

Efectuarea unor calcule simple ne indică şi de această dată că, de fapt, volumul eşantionului obţinut nu reflectă variaţiile mărimii colectivităţii totale. Se constituie anumite praguri peste care **n** (volumul eşantionului) nu mai creşte oricât de mult ar creşte **N**.

Aşa, de exemplu, la Δ_x = 5% pragul respectiv este 399. La valori mai mici ale erorii limită admise pragul eşantionului se fixează, evident, la valori mai mari, dar întotdeauna în jurul valorii indicate de expresia de definiţie a mărimii eşantionului în care **P = 50%** (când furnizează cea mai mare mărime a eşantionului), iar **t = 1,96** corespunde unei probabilităţi de **95%**.

După cum se observă, metoda Jamane poate conduce la eşantioane subdimensionate (când se doreşte o siguranţă mai mare, deci o probabilitate de peste 95%, sau când populaţia este eterogenă în raport cu caracteristicile de bază) şi la eşantioane supradimensionate (în cazul în care populaţia este relativ omogenă). Şi de această dată dorinţa de a lucra cu erori mici de reprezentativitate conduce la eşantioane foarte mari, tabelul 4.2.

Material cu drept de autor

Tabelul 4.2

Mărimile simulate ale eşantionului calculat cu expresia (4.12) pentru anumite valori ale lui **N** şi Δ_x

Mărimea colectivității totale	Mărimea eşantionului pentru diverse erori limită admise		
	Δ_x **= 5%**	Δ_x **= 3%**	Δ_x **= 1%**
500	222	345	476
1000	285	526	909
5000	370	909	3333
10000	384	1000	5000
100000	398	1099	9090
1000000	399	1109	9900
1000000000	399	1110	9990

Calculul expus îndreptățește, pe de o parte, afirmația potrivit căreia mărimea eşantionului nu depinde în mod absolut de mărimea colectivității totale (în condițiile în care eroarea de reprezentativitate rămâne constantă, variațiile mărimii colectivității totale nu pot fi reflectate în mărimea eşantionului), iar, pe de altă parte, impune o anumită circumspecție în utilizarea metodei „simplificate" pentru determinarea volumului eşantionului.

Stabilirea volumului eşantionului se face pe bază unui compromis între opțiunea pentru eşantioane de volum mare, care să asigure un grad mare de reprezentativitate şi un grad mare de încredere pentru parametrii estimați ai populației de referință, şi opțiunea pentru eşantioane de volum mic, ce implică costuri reduse. Dimensiunea minimă a eşantionului trebuie să asigure o reprezentativitate acceptabilă în procesul inferenței statistice, care să nu ducă la distorsiuni.

Mărimea eşantionului depinde de numeroşi factori controlabili şi necontrolabili pentru cercetător: gradul de exactitate cu care se doreşte să se estimeze caracteristicile populației de referință, mărimea erorilor de sondaj, legea numerelor mari şi, nu în ultimul rând, bugetul disponibil, perioada de timp avută la dispoziție şi resursele de personal de care dispune.

Gradul preciziei cerute de beneficiarul rezultatelor este principalul factor ce determină mărimea eşantionului. Încrederea ce poate fi atribuită informațiilor obținute pe baza unui eşantion depinde direct de mărimea eşantionului, şi nu de fracția de selecție.

În stabilirea dimensiunii eşantionului, dacă se stabileşte aprioric un prag de semnificație α, se impune ca abaterile dintre media populației de referință, dacă se cunosc, şi mediile eşantioanelor ce se pot genera să respecte inegalitatea.

Gradul de variabilitate al oricărei populații este un alt factor de influență ce poate fi sau nu cunoscut. Dacă variabilitatea este cunoscută din

Material cu drept de autor

cercetări anterioare sau dintr-o cercetare organizată în mod special, volumul eşantionului rezultă imediat din calcule. Dacă dispersia eşantionului sau a populației de referință nu este cunoscută în cazul cel mai nefavorabil), se ia în calcul dispersia maximă. Calculul ei presupune stabilirea mărimii maxime a dispersiei pentru caracteristici cantitative:

$$\sigma^2_{\max} = \frac{(x_{\min} - \bar{x})^2 + (x_{\max} - \bar{x})^2}{2} \quad şi \qquad (4.13)$$

$$\sigma^2_{\max} = f.(1-f) = 0{,}5.(1-0{,}5) = 0{,}25$$

şi determinarea valorii de 0,25 – corespunzătoare frecvenței maxime **f,** a caracteristicii alternative.

În practică se operează cu eşantioane de volum redus (pentru care se foloseşte la estimarea erorilor legea repartiție Student) şi eşantioane de volum normal (pentru care se foloseşte legea de repartiție Laplace), în funcție de gradul de omogenitate al colectivității. Evident, conform legii numerelor mari cu cât creşte volumul eşantionului (nu fracția de selecție), cu atât precizia rezultatelor este mai mare. Mărimea eşantionului se decide şi în funcție de rezultatele ce vor fi analizate, ținându-se seama de necesitatea obținerii preciziei nu doar pe total eşantion, ci şi pe subgrupe.

Un alt factor de influență este faptul că sondajul, în general, urmăreşte rezultatele privitoare la mai multe caracteristici. Un eşantion suficient de mare pentru estimarea intervalului de încredere pentru o caracteristică poate să fie insuficient pentru o altă caracteristică.

Problemele se complică în cazul sondajelor stratificate, caz în care trebuie estimată mărimea eşantionului pe fiecare strat şi apoi, prin însumare, va rezulta eşantionul pe total populație. În cazul sondajului multistadial trebuie făcută o estimare a variantelor în interiorul şi între unitățile din primul stadiu. Toate aceste dificultăți de natură tehnică pot fi evitate printr-o documentare prealabilă corespunzătoare.

Mărimea eşantionului va mai fi decisă şi în funcție de restricțiile de resurse şi bugetul disponibil, de cerințele beneficiarului şi de posibilitatea previzionării dispersiei variabilelor înregistrate.

4.5. Probleme privind prognoza volumului de eşantionare. Dispersii marginale

Fenomenele din domeniul studiului fenomenelor economico-sociale sunt, în general, fenomene dinamice, ceea ce determină valori diferite în timp ale variabilelor ce le caracterizează. Sondajul prezintă o situație statică, de aceea se recomandă organizarea de sondaje periodice cu acelaşi set de variabile înregistrate.

Măsurând valorile unor variabile ce caracterizează un fenomen în momente diferite de timp putem determina variații statistice nu doar între valorile individuale înregistrate, ci şi între valorile medii calculate la momente de timp diferite, păstrându-se sau nu aceeaşi amplitudine a variației.

Material cu drept de autor

Pentru proiectarea volumului unui nou eşantion în sondaje periodice este necesară cunoaşterea tendinței de evoluție a dispersiei şi posibilitatea previzionării dispersiei şi abaterii standard. De aici apare ca necesară elaborarea de serii de timp de dispersii şi de modificări absolute şi/sau relative ale acesteia, ca de exemplu, serii cronologice de indici ai dispersiei sau de sporuri ale dispersiei unei variabile înregistrate în sondaje efectuate la momente de timp diferite.

Cea mai frecventă metodă de sondaj utilizată în studiile sociale şi economice este sondajul stratificat. De aceea, această formă de sondaj poate fi utilizată şi pentru studiul fenomenelor în dinamică, pe baza datelor înregistrate în sondaje realizate în perioade diferite, în vederea elaborării de serii cronologice care să permită, prin aplicarea de tehnici complexe de sondaj, efectuarea de prognoze optime.

În scopul previzionării gradului de variație, a dispersiei şi a abaterii tip, în urma realizării de sondaje periodice se pot construi serii cronologice cu periodicitate constantă sau variabilă, de niveluri atinse de indicatori micro sau macroeconomici, de medii şi măsuri ale variației atinse de valorile individuale ale distribuției marginale şi chiar de indicatori marginali care exprimă modificarea mediilor şi a dispersiilor.

Cunoaştem că rezultatele unui sondaj stratificat conduc la verificarea regulii de adunare a dispersiilor, conform căreia dispersia totală înregistrată de variația valorilor individuale ale distribuției marginale este suma dispersiilor parțiale, deci:

$$\sigma^2_{total} = \overline{\sigma}^2 + \sigma^2_{y/x} \qquad (4.14)$$

unde: σ^2_{total} - dispersia totală, determinată de toți factorii de influență ai variației unei variabile; $\overline{\sigma}^2$ - media dispersiilor înregistrată în interiorul straturilor determinată de factorii neînregistrați; $\sigma^2_{y/x}$ - dispersia dintre straturi, determinată de factorul de formare a straturilor şi arată în ce măsură discriminează sau nu criteriul de stratificare variabila studiată.

Dacă simplificăm fiecare termen al ecuației de mai sus cu dispersia totală, calculând deci structura dispersiei totale, obținem raportul de determinație şi raportul de nedeterminație, după formula:

$$1 = \frac{\overline{\sigma}^2}{\sigma^2_{total}} + \frac{\sigma^2_{y/x}}{\sigma^2_{total}} \qquad (4.15)$$

unde $\overline{\sigma}^2 / \sigma^2_{total}$ - raport de nedeterminație ce exprimă procentual partea din varianța totală datorată factorilor aleatori neînregistrați; $\sigma^2_{y/x} / \sigma^2_{total}$ - raport de determinație ce exprimă procentual partea din varianța totală a variabilei dependente explicată de factorul de grupare, de discriminare, de variabila independentă.

Material cu drept de autor

Dacă înregistrăm două niveluri ale dispersiei pentru două sondaje succesive se poate calcula modificarea absolută a dispersiei totale ce se va distribui între modificarea absolută a dispersiei dintre straturi şi a mediei dispersiilor din interiorul straturilor, astfel:

$$\Delta\sigma^2_{total} = \Delta\overline{\sigma}^2 + \Delta\sigma^2_{y/x} \qquad (4.16)$$

Acest spor al dispersiei totale poate fi pozitiv sau negativ şi se poate distribui egal sau diferit pe cele două componente. Pentru a măsura modul de distribuție şi a determina contribuția factorului de stratificare la variația dispersiei totale, calculăm structura ecuației modificărilor absolute ale dispersiilor împărțind ecuația modificărilor absolute, cu modificarea absolută a dispersiei totale, astfel:

$$1 = \frac{\Delta\overline{\sigma}^2}{\Delta\sigma^2_{total}} + \frac{\Delta\sigma^2_{y/x}}{\Delta\sigma^2_{total}} \qquad (4.17)$$

În ecuația (4.17) se propune ca indicatorii marginali obținuți să se noteze, numească şi interpreteze astfel:

a. $\overline{\sigma}^2_{mg} = \Delta\overline{\sigma}^2 \big/ \Delta\sigma^2_{total}$ = dispersie medie marginală, ce arată cu cât se va modifica media dispersiilor din interiorul straturilor dacă dispersia totală suferă o modificare cu o unitate sau cu cât trebuie să se modifice media dispersiilor din interiorul straturilor pentru a obține o modificare cu o unitate a dispersiei totale; poate lua valori între -1 şi +1.

b. $\sigma^2_{y/x\,mg} = \Delta\sigma^2_{y/x} \big/ \Delta\sigma^2_{total}$ = dispersie marginală dintre straturi, ce arată cu cât se va modifica nivelul dispersiei dintre straturi sau cu cât trebuie modificată dispersia dintre straturi pentru a obține o modificare unitară a dispersiei totale; poate lua valori între -1 şi +1.

Există posibilitatea identificării unei funcții matematice de trend pe termen lung, atât a dispersiilor marginale, cât şi a raportului acestora cu evoluția în domeniu. Între cele trei tipuri de dispersii există o relație directă sau inversă, deci creşterea dispersiei totale va determina creşteri/descreşteri în proporții egale sau diferite ale dispersiilor parțiale, şi invers.

Desigur dispersiile de eşantion sunt corectate cu numărul gradelor de libertate corespunzătoare, dar pentru simplificarea modului de scriere a formulelor nu am mai introdus şi aceste notații.

Dispersia totală se corectează cu **n - 1** grade de libertate, dispersia dintre starturi se corectează cu numărul de straturi, deci **r - 1**, iar media dispersiilor din interiorul starturilor cu volumul eşantionului, adică numărul de straturi, deci **n - r**.

Construind serii cronologice de dispersii marginale ce vor fi supuse analizei statistice de previziune putem estima, cu o anumită probabilitate, nivelul mediei dispersiei din interiorul straturilor şi al dispersiei totale, niveluri necesare programării unui nou volum de eşantionare. Dacă seriile construite sunt nestaționare vor trebui diferențiate pentru a se transforma în evoluții staționare.

Material cu drept de autor

În final, trebuie precizat că modificările absolute ale dispersiilor corectate pot fi calculate cu bază mobilă sau cu bază fixă. Sporurile cu bază fixă apar în cazul în care am realizat într-o cercetare anterioară o probă martor sau un eşantion programat în care se ajunge la o distribuție martor ce coincide cu structura distribuției totale şi a cărei reprezentativitate este validată statistic.

Necesitatea utilizării indicatorilor marginali ai variației valorilor individuale ale unei variabile cantitative de sondaj este legată, mai ales, de calculele de prognoză care sunt necesare pentru determinarea volumului unui nou eşantion.

Metoda se poate aplica în special în situația sondajului stratificat, caz în care, pentru estimarea intervalului de încredere, se foloseşte media dispersiilor din interiorul straturilor şi necesită, pentru o mai bună fundamentare teoretică, testarea riguroasă în activitatea practică.

4.6. Determinarea volumului eşantionului pentru selecția aleatoare simplă

Practica sondajului demonstrează că selecția aleatoare simplă poate fi folosită cu succes numai în studierea unor colectivități monotipice care prezintă un grad ridicat de omogenitate. În acest caz, eşantionul se formează din unități simple care se extrag din colectivitatea generală prin procedeul repetat sau nerepetat pe baza unei scheme probabiliste.

Acest tip de selecție dispune de cele mai simple formule de calcul a indicatorilor de selecție care cu unele modificări se folosesc ca bază de calcul şi în celelalte tipuri de selecție. Se poate spune pe drept cuvânt că deşi acest tip de selecție nu dă rezultate bune în cazul colectivităților neomogene, prezintă avantajul că multe din principiile fundamentale ale selecției pot fi explicate pe baza selecției aleatoare simple. În această situație sondajul simplu apare ca o variantă a procedeului bilei revenite sau nerevenite.

Intervalul de încredere al mediei colectivității generale este determinat de mărimea medie a eşantionului şi de eroarea limită respectivă, astfel:

- pentru caracteristica nealternativă:

$$\bar{x}_s - \Delta_x < \bar{x}_0 < \bar{x}_s + \Delta_x$$

- pentru caracteristica alternativă:

$$w - \Delta_w < p < w + \Delta_w$$

În mod asemănător se poate calcula intervalul de încredere al oricărui alt parametru.

Determinarea intervalului de variație al mediei estimate pe baza datelor de selecție permite şi stabilirea intervalului de variație al nivelului totalizat al caracteristicii care se poate determina după relația:

- pentru caracteristica nealternativă:

Material cu drept de autor

$$N.(\bar{x}_s - \Delta_x) < \sum_{i=1}^{N} x_i < N.(\bar{x}_s + \Delta_x)$$

- pentru caracteristica alternativă:

$$N.(w - \Delta_w) < M < N.(w + \Delta_w)$$

După cum s-a arătat precizia rezultatelor selecției, posibilitatea extinderii lor asupra întregii colectivități, depinde şi de numărul de unități la care se face culegerea datelor.

Dacă se consideră, drept criteriu de eficiență a sondajului mărimea erorii de estimare a mediei generale, atunci trebuie ca volumul eşantionului să corespundă relației:

$$z_\alpha . \sigma_{\bar{x}} \leq \Delta_{\bar{x}}$$

Pentru determinarea limitei minime a volumului eşantionului se ține seama de procedeul de selecție aplicat. Pentru sondajul simplu repetat va fi:

$$\Delta_{\bar{x}} = z_\alpha \cdot \sqrt{\frac{\sigma_0^2}{n}}$$

de unde:

$$n = \frac{z_\alpha^2 . \sigma_0^2}{\Delta_x^2} \qquad (4.18)$$

Această mărime este minimă pentru asigurarea gradului de reprezentativitate dorit.

Pentru sondajul simplu nerepetat:

$$\Delta_{\bar{x}} = z_\alpha \cdot \sqrt{\frac{\sigma_0^2}{n}\left(1 - \frac{n}{N}\right)}$$

de unde:

$$n = \frac{z_\alpha^2 . \sigma_0^2}{\Delta_x^2 + \frac{z_\alpha^2 . \sigma_0^2}{N}} \qquad (4.19)$$

Comparând cele două relații rezultă că pentru acelaşi grad de reprezentativitate volumul eşantionului este cel mai mic dacă se foloseşte procedeul bilei nerevenite.

Material cu drept de autor

Exemplul de calcul 4.1. Să se determine numărul de piese ce trebuie extrase, în mod aleatoriu şi nerepetat, dintr-un lot de **3000** piese dacă diametrul mediu al pieselor din eşantion trebuie să difere de diametrul mediu al pieselor din întreaga populație statistică cu cel mult 0,2 mm.

Dintr-o cercetare anterioară se cunoaşte că varianța caracteristicii studiate (diametrul mediu) a fost de 20 mm. Rezultatul se garantează cu o probabilitate de **0,9962** pentru care **z = 2,90**.

Rezolvare. Se cunosc: **N = 3000;** Δ_x **= 0,2 mm;** σ^2 **= 0,2 mm; z = 2,90**. Se cere să se determine **n = ?**.

Se utilizează relația (4.11) din manual:

$$n = \frac{z^2.\sigma^2}{\Delta_x^2 + \frac{z^2.\sigma^2}{N}} = \frac{2{,}9^2.0{,}2}{0{,}04 + \frac{2{,}9^2.0{,}2}{3000}} = \frac{8{,}41.0{,}2}{0{,}04 + \frac{8{,}41.0{,}2}{3000}} =$$

$$= \frac{1{,}682}{0{,}04 + 0{,}00056} = \frac{1{,}682}{0{,}0456} = 36{,}89 \approx 37$$

Rezultă că pentru a evalua diametrul mediu al pieselor din întreaga populație statistică de 3000 de piese, în limitele precizate de problemă, volumul eşantionului va trebui să fie format de circa 42 de piese.

Exemplul de calcul 4.2. În condițiile datelor problemei precedente, se cere să se rezolve problema în ipoteza că sondajul s-a efectuat aleatoriu şi repetat.

Rezolvare. Se cunosc: **N = 3000;** Δ_x **= 0,2 mm;** σ^2 **= 0,2 mm; z = 2,90**. Se cere să se determine **n = ?**. Folosindu-se relația (4.10) se obține:

$$n = \frac{z^2.\sigma^2}{\Delta_x^2} = \frac{2{,}9^2.0{,}2}{0{,}04} = \frac{8{,}41.0{,}2}{0{,}04} = \frac{1{,}682}{0{,}04} = 42{,}05 \approx 42$$

După cum se observă, în cazul sondajului repetat, volumul eşantionului rezultă ceva mai mare decât în cazul sondajului aleatoriu nerepetat.

Exemplul de calcul 4.3. La o stațiune montană de odihnă şi tratament, într-o lună, s-au înregistrat în total **3200** turişti, pe sexe: **1376** masculin şi **1824** feminin.

Se cere:

1. să se stabilească volumul necesar al eşantionului folosind drept caracteristică de reprezentativitate repartiția pe sexe **P = 0,95** (**z = 1,96**), iar eroarea maximă admisă de **5%**;

2. efectuând un sondaj pilot şi prelucrând datele din eşantion s-a obținut o durată medie a sejurului de 10 zile/turist cu o abatere de 4 zile; să se dimensioneze un nou eşantion garantând rezultatele cu o probabilitate P = 0,9973 (z = 3).

Material cu drept de autor

Rezolvare:

1. Folosind drept caracteristică de asigurare a reprezentativității o variabilă alternativă (sexul), se va obține:

$$n = \frac{z^2.p.(1-p)}{\Delta_x^2 + \frac{z^2.p.(1-p)}{N}} = \frac{1{,}96^2.0{,}43.(1-0{,}43)}{0{,}05^2 + \frac{1{,}96^2.0{,}43.(1-0{,}43)}{3200}} \cong 337 \ turisti$$

În expresia de calcul: **p = M/N = 1376/3200 = 0,43**.

În practică se consideră eroarea limită de **3%** în care caz: **n = 789** turişti.

2. Întrucât durata sejurului este variabilă numerică, eroarea limită se poate determina astfel:

Δ_x **= (5/100).10 = 0,5**.

Volumul noului eşantion va fi:

$$n' = \frac{z^2.\sigma^2}{\Delta_x^2 + \frac{z^2.\sigma^2}{N}} = \frac{1{,}96^2.16}{0{,}25 + \frac{1{,}96^2.16}{3200}} = 121 \ turisti$$

Exemplul de calcul 4.4. Un post de televiziune doreşte să organizeze o eşantionare aleatoare simplă nerepetată pentru a estima procentul locuințelor dintr-o localitate care urmăresc ştirile de la ora 19,00. Care este volumul necesar al eşantionului pentru o eroare de 4% şi o probabilitate de garantare a rezultatelor de 95% ?

Rezolvare.

Deşi nu se cunoaşte dispersia populației, se ştie totuşi împrăştierea maximă care este: w.(1 – w) = 0,5x0,5 = 0,25. Atunci:

$$\Delta_w = z.\sigma_w = z.\sqrt{\frac{w.(1-w)}{n}\left(1-\frac{n}{N}\right)}; \qquad z_{0,025} = 1{,}96$$

Cum numărul locuințelor din localitate (N) este considerat foarte mare se poate considera coeficientul de corecție (1 – n/N) ≈ 1. Deci:

$$n = \frac{z^2.w.(1-w)}{\Delta_w^2} = \frac{1{,}96^2.0{,}5.0{,}5}{(0{,}04)^2} = 600 \ persoane$$

Material cu drept de autor

Cap.5. PRELUCRAREA DATELOR STATISTICE

5.1.Analiza preliminară a datelor statistice

După culegerea datelor experimentale şi ordonarea lor tabelară se procedează la analiza preliminară a lor stabilindu-se corectitudinea lor, din punctul de vedere al reprezentativității calculelor statistice ulterioare. Ne aflăm aşa dar în fața unei baze de date care conține rezultatele cercetării şi cel mai puternic impuls este acela de a trece cât mai rapid la prelucrarea lor. Oricât de motivați am fi să finalizăm cât mai repede prelucrarea datelor, avem motive serioase să le supunem unui analize preliminare. În absența unei riguroase analize de acest fel, riscurile obținerii unor rezultate viciate şi, pe această bază, riscul unor concluzii greşite, este foarte mare.

Dintre cele mai importante aspecte care fac obiectul analizei preliminare, următoarele ni se par a fi absolut indispensabile:

1. Corectitudinea datelor. Calitatea datelor de cercetare face, desigur, obiectul unei atenții susținute încă din faza proiectării şi aplicării instrumentelor şi a recoltării acestora. Acum ne vom referi, însă, la verificarea finală, care trebuie efectuată înainte de a trece efectiv la prelucrarea datelor, atunci când datele se află deja introduse într-o bază de date computerizată.

Ideală ar fi o corectură minuțioasă a tuturor valorilor. Rutina şi oboseala inerentă pot conduce adesea la erori de introducere a datelor. Acest lucru este relativ uşor de făcut cu date puține, dar descurajant atunci când avem un număr mare de valori. De aceea este recomandabil ca în faza de introducere a datelor să se adopte măsuri de prevenire şi de corecție operativă a erorilor (evitarea lucrului sub presiune, pauze pentru odihnă, corecții parțiale, introducerea datelor în echipă).

a. Situația este ceva mai bună atunci când datele primare rezultă din aplicarea computerizată a instrumentelor de cercetare, cu înregistrarea directă a rezultatelor. Chiar şi în acest caz, însă, se mențin surse de imprecizie şi de erori, care determină înregistrarea unor valori atipice, care merită atenție.

2. Valorile excesive (marginale şi extreme). Nu este de loc rară situația în care valorile problematice ale unei distribuții se află la extremele acesteia, în zona valorilor cele mai mari sau a celor mai mici. Această situație poate proveni din mai multe surse.

a. Erori de tastare la introducere (de exemplu: „422" în loc de „42")

b. Valori corect măsurate şi înregistrate dar care exprimă o altă realitate decât cea pe care am dori să o măsurăm (de exemplu: o valoare aberantă a timpului de reacție, determinată de distragerea conjuncturală a atenției).

c. Valori care exprimă realmente o caracteristică a subiecților respectivi dar care fac parte dintr-o altă categorie decât ceilalți subiecți din eşantion.

Material cu drept de autor

3. Valorile lipsă (*missing values*). Într-o cercetare, valorile lipsă sunt informații care nu au putut fi recoltate din diverse motive (subiecți care au refuzat sau au uitat să completeze anumite date, imposibilitatea măsurării, etc.). În general, lipsa unora dintre date nu afectează în mod semnificativ analiza statistică, mai ales dacă numărul cazurilor valide este suficient de mare. Totuşi, o analiză a datelor lipsă este necesară pentru a evalua amploarea fenomenului şi a decide dacă acesta este unul ocazional, aleator, putând fi ignorat, sau are un caracter sistematic, fapt care ar putea afecta concluziile cercetării.

Să ne imaginăm că în cazul unui studiu cu privire la satisfacția în muncă într-o bancă, se constată că există un număr de subiecți care nu au completat informația cu privire la serviciul în care sunt angajați. Putem presupune că acest fapt se datorează neatenției în completarea datelor sau, dimpotrivă, că subiecții respectivi provin din servicii în care există probleme de conducere. Într-un asemenea caz, semnificația valorilor lipsă se poate testa cu ajutorul unei proceduri simple:

- Se creează o variabilă de lucru, care ia două valori, să zicem „0" pentru subiecții care nu au răspuns şi „1" pentru cei care au răspuns la întrebarea respectivă.
- Se aplică testul *t* al diferenței dintre medii pentru „satisfacția în muncă, pe cele două categorii de subiecți.
- O valoare semnificativă ar conduce la concluzia că valorile lipsă sunt efectul unei reacții de apărare a subiecților în timp ce o valoare nesemnificativă a testului ***t*** ar sugera faptul că valorile lipsă au un caracter nesistematic.

Dacă se consideră că valorile lipsă nu pot fi ignorate, există mai multe posibilități de intervenție. Prima, şi cea mai simplă, este aceea de a elimina cazurile cu valori lipsă. Dacă valorile lipsă aparțin cu precădere anumitor variabile, aceste variabile ar putea fi eliminate din analiză. În ambele situații există riscul diminuării cazurilor valide. Atunci când studiul urmăreşte efectuarea unor comparații între subgrupe de subiecți, se poate ajunge la situația în care unele dintre acestea să se reducă numeric până la imposibilitatea efectuării analizelor respective.

O altă soluție este aceea de a înlocui valorile lipsă. Cercetătorul are posibilitatea să introducă valori, atunci când experiența şi cunoaşterea domeniului îi permit să le evalueze. După alegerea variabilei pentru care se aplică procedura de înlocuire a valorilor lipsă, urmează alegerea metodei de înlocuire, care poate utiliza una dintre următoarele valori:

- media variabilei pe întreaga distribuție;
- media valorilor valide din vecinătatea valorilor lipsă;
- mediana valorilor valide din vecinătatea valorilor lipsă;
- interpolarea liniară între valoarea validă anterioară şi cea posterioară valorii lipsă;
- tendința liniară a punctului reprezentat de valoarea lipsă (predicție liniară a valorilor lipsă).

Material cu drept de autor

Atunci când se alege soluția înlocuirii valorilor lipsă se recomandă efectuarea prelucrărilor atât cu valorile „recuperate" cât şi cu ele lipsă. Dacă rezultatele sunt similare, se vor accepta transformările, în caz contrar, cercetătorul trebuie să decidă care din cele două prelucrări este adecvată situației respective.

4. Normalitatea distribuției. Testele statistice parametrice se bazează pe asumarea unor condiții esențiale. Printre acestea, normalitatea distribuției variabilei dependente este cea mai importantă. De aceea, verificarea normalității este una dintre problemele de neocolit în faza de analiză primară a datelor de cercetare.

Normalitatea poate fi apreciată empiric cu ajutorul unei proceduri de reprezentare grafică. Histograma permite suprapunerea curbei normale teoretice corespunzătoare parametrilor reali ai distribuției respective (media şi abaterea standard), peste distribuția reală.

5.2. Criterii pentru eliminarea valorilor ce diferă semnificativ de restul selecției

Dacă se constată o neomogenitate a datelor statistice, ca rezultat al acțiunii unor cauze accidentale, este necesară excluderea valorilor ce reprezintă abateri grosolane de la şirul valorilor observate, cu alte cuvinte este necesară omogenizarea materialului statistic obținut în cercetare. Această operație se efectuează prin intermediul criteriilor pentru eliminarea valorilor însoțite de erori grosolane. Există mai multe criterii pentru eliminarea valorilor ce diferă semnificativ de la omogenitatea acestora. Vom menționa unele dintre ele.

1. Testul Q. este cel mai indicat pentru selecțiile foarte mici (principal pentru **n < 4**). Are expresia:

$$Q = \frac{x_1 - x_2}{R} \qquad (5.1)$$

unde $\mathbf{x_1}$ este valoarea nesigură, $\mathbf{x_2}$ este valoarea cea mai apropiată de aceasta, iar **R** este amplitudinea.

Valoarea Q calculată cu relație (5.1) se compară cu valoarea Q tabelată în funcție de probabilitatea P% şi numărul de determinări n, tabelul 5.1.

Tabelul 5.1

Valori critice pentru criteriul Q

n	P=90%	P=95%	P=99%
3	0,89	0,94	0,99
4	0,68	0,77	0,89
5	0,56	0,64	0,76
6	0,48	0,56	0,70
7	0,43	0,51	0,64
8	0,40	0,48	0,58

Material cu drept de autor

Exemplu de calcul 5.1. La determinarea repetată a procentului de carbon din fontă s-au găsit valorile: **2,86; 2,89; 2,90; 2,91; 2,99**. Se întreabă dacă valoarea **2,99** este nesigură ?

$$Q = \frac{2,99 - 2,91}{2,99 - 2,86} = 0,62$$

În tabelul 5.5 se găseşte **Q(P=95%** şi **n=5) = 0,64**. Deoarece $\mathbf{Q_{calc} < Q_{tab}}$ nu există nici un temei (în cadrul a 5 determinări) de a privi valoarea **2,99** ca anormală.

2. Testul lui Graf şi Henning. Este recomandat pentru selecțiile cu **n > 4**.

Intervalul de toleranță este dat de expresia:

$$\bar{x} \pm g(P,n)s \qquad (5.2)$$

în care $\bar{\mathbf{X}}$ este media aritmetică, iar **S** este abaterea medie pătratică calculată pe baza tuturor valorilor, inclusiv valoarea dubioasă. Mărimea **g(P,n)** se stabileşte cu ajutorul nomogramei prezentată în figura 5.1.

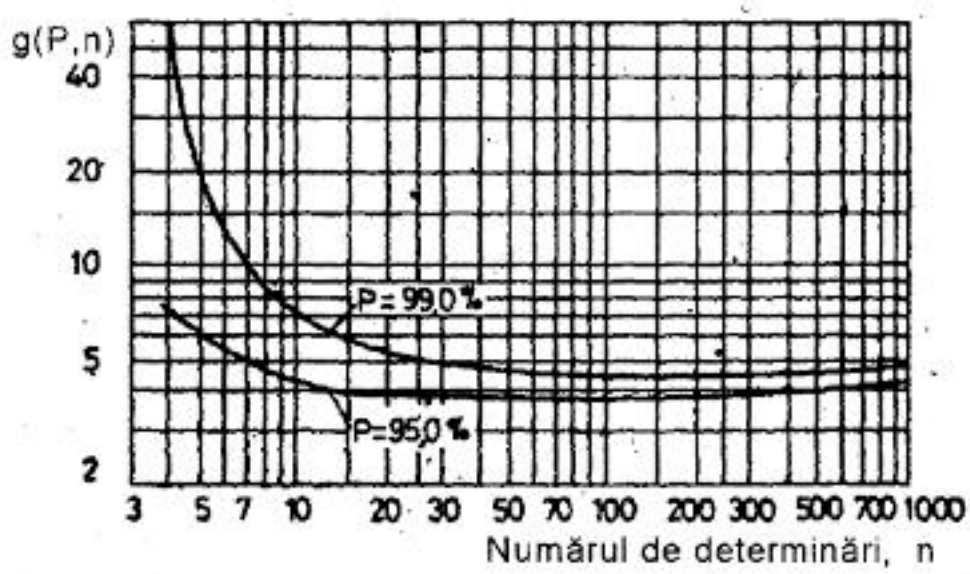


Fig.5.1. Nomograma pentru eliminarea rezultatelor anormale

Exemplu de calcul 5.2. Să se stabilească dacă valoarea **2,99** din exemplul de calcul 5 este sau nu normală ? Parametrii statistici calculați sunt: $\bar{\mathbf{x}} = \mathbf{2,91}$ şi $\mathbf{S = 0,049}$.

Mărimea **g(P=95%,n=5) = 5,9**. Astfel avem: **2,91±5,9x0,49 = 2,91±0,29** este o valoare nesigură; **2,99**, găsindu-se în interiorul intervalului de toleranță nu este de fapt anormală.

3. Testul t. Este justificat de faptul că la baza procedeelor de testare a rezultatelor dubioase în cazul selecțiilor mici stă, în general, repartiția **t** a lui Student. Pentru eliminarea rezultatelor anormale, testul **t** se foloseşte în felul următor. Se calculează mărimea **t**:

$$t = \frac{\bar{x}_{n-1} - x_d}{s\sqrt{n/(n-1)}} \qquad (5.3)$$

unde $\bar{x}_{n-1}$ este media aritmetică calculată pentru determinările rămase după eliminarea valorii dubioase $\mathbf{x_d}$, respectiv a **n-a** determinare. Abaterea standard, ce se calculează fără valoarea $\mathbf{x_d}$, este:

Material cu drept de autor

$$s = \sqrt{\frac{\sum \left(x_i - \bar{x}_{n-1}\right)^2}{n-2}}$$

Expresia (5.19) rezultă din expresia testului **t** pentru compararea a două medii, în care drept cea de a doua medie se ia o valoare individuală, respectiv $\mathbf{x_d}$, ceea ce atrage după sine şi $\mathbf{n_1 = n-1;\ n_2 = 1}$. Dacă **t** calculat pe baza relației (5.3) este mai mare decât $\mathbf{t_p}$ tabelat, la nivelul de încredere ales, pentru **n-2** grade de libertate, înseamnă că valoarea $\mathbf{x_d}$ este dubioasă. Dacă $\mathbf{t_{calculat} < t_{tabelat}}$, valoarea $\mathbf{x_d}$ nu este anormală şi se păstrează în selecție.

Exemplu de calcul, 5.3. Se dau următoarele 10 valori medii de selecție: **0,26; 0,21; 0,20; 0,21; 0,21; 0,19; 0,18; 0,17; 0,18; 0,19**. Se cere să se stabilească dacă $\mathbf{x_d = 0,26}$ este o valoare anormală. Se obțin: $\bar{x}_{n-1} = 0,193$ şi $s = 0,015$. Pe baza relației (5.19) avem:

$$t = \frac{0,193 - 0,26}{0,015\sqrt{10/9}} = 4,24$$

În anexa 3 găsim $\mathbf{t_P}$ **(P=95%; q/2=2,5%; v=n-2=8)=2,306**. Întrucât $\mathbf{t = 4,24 < t_P = 2,306}$ înseamnă că valoarea **0,26** este anormală.

Considerând în continuare că eventual valoarea 0,17 poate fi dubioasă, se utilizează din nou testul t şi se obține $\mathbf{\bar{x}_8 = 0,146; S = 0,0134; t = 1,83}$**. În anexa 3 găsim** $\mathbf{t_P}$ **(P=95%; q/2=2,5%; ν = 8) = 2,37. Întrucât** $\mathbf{t = 1,83 < t_P = 2,37}$ **valoarea 0,17 se păstrează în selecție.**

4. Testul Irvin sau λ_P. Pentru aplicarea testului Irvin, valorile înregistrate pentru variabila aleatoare se aranjează în ordine crescătoare:

$$x_1 \le x_2 \le x_3 \le \ldots \le x_{n-2} \le x_{n-1} \le x_n$$

Tabelul 5.2

Valori admisibile pentru λ_P

α	Valori limită pentru λ_P în funcție de numărul de probe elementare **n** şi nivelul de semnificație α ales								
	2	3	10	20	30	50	100	400	1000
0,10	2,3	1,8	1,2	1,0	1,0	0,9	0,8	0,7	0,6
0,05	2,8	2,2	1,5	1,3	1,2	1,1	1,0	0,9	0,8
0,01	3,7	2,9	2,0	1,8	1,7	1,6	1,5	1,3	1,2
0,005	4,0	3,2	2,3	2,0	1,9	1,8	1,6	1,5	1,4

Se determină abaterea medie pătratică de selecție **s** şi apoi se determină mărimile:

$$\lambda_1 = \frac{x_2 - x_1}{s} \qquad (5.4)$$

Material cu drept de autor

respectiv:

$$\lambda_n = \frac{x_n - x_{n-1}}{s} \tag{5.5}$$

Dacă $\lambda_1 \leq \lambda_P$ sau $\lambda_n \leq \lambda_P$, valorile admisibile pentru λ_P fiind prezentate în tabelul 5.2, nu există nici un temei să se considere că şirul de valori întâmplătoare nu este omogen.

Dacă $\lambda_1 > \lambda_P$ sau/şi $\lambda_n > \lambda_P$, atunci variabilele $\mathbf{x_1}$ sau/şi $\mathbf{x_n}$ se elimină din şirul variabilei aleatoare înregistrate, se reface calculul pentru abaterea medie pătratică **s** corespunzătoare noului şir de variabile şi se refac, în noile condiţii, calculele respective, până când se obţine omogenitatea datelor înregistrate în şirul variabilei aleatoare X.

5. Testul Grubbs. Ca şi la testul Irvin, datele înregistrate pentru şirul de valori întâmplătoare se ordonează în ordine crescătoare, se determină valoarea medie $\bar{x}$ şi abaterea medie pătratică **s** pentru şirul de valori şi apoi se calculează mărimile:

$$v_1 = \frac{\bar{x} - x_1}{s} \tag{5.6}$$

respectiv:

$$v_n = \frac{x_n - \bar{x}}{s} \tag{5.7}$$

Tabelul 5.3

Valori critice pentru testul Grubbs

n	Valoarea v_P pentru α şi n				n	Valoarea v_P pentru α şi n			
	0,10	0,05	0,025	0,01		0,10	0,05	0,025	0,01
3	1,406	1,412	1,414	1,414	15	2,326	2,493	2,638	2,800
4	1,645	1,689	1,710	1,723	16	2,354	2,523	2,670	2,837
5	1,791	1,869	1,917	1,955	17	2,380	2,551	2,701	2,871
6	1,894	1,996	2,967	2,130	18	2,404	2,577	2,728	2,903
7	1,974	2,093	2,182	2,265	19	2,426	2,600	2,754	2,932
8	2,041	2,172	2,273	2,374	20	2,447	2,623	2,778	2,959
9	2,097	2,237	2,349	2,464	21	2,467	2,644	2,801	2,984
10	2,146	2,294	2,414	2,540	22	2,486	2,664	2,823	3,008
11	2,190	2,343	2,470	2,606	23	2,504	2,683	2,843	3,030
12	2,229	2,387	2,519	2,663	24	2,520	2,701	2,862	3,051
13	2,264	2,426	2,562	2,714	25	2,537	2,717	2,880	3,071
14	2,297	2,461	2,602	2,759					

Material cu drept de autor

În tabelul 5.3 se prezintă valorile critice pentru v_P în funcție de volumul eşantionului **n** şi nivelul de semnificație ales α .

Dacă $v_1 \leq v_P$ sau/şi $v_n \leq v_P$, valorile admisibile pentru v_P fiind prezentate în tabelul 5.7, nu există nici un temei să se considere că şirul de valori întâmplătoare nu este omogen.

Dacă $v_1 > v_P$ sau/şi $v_n > v_P$, atunci variabilele x_1 sau/şi x_n se elimină din şirul variabilei aleatoare înregistrate, se reface calculul pentru valoarea medie $\bar{x}$ şi abaterea medie pătratică **s** corespunzătoare noului şir de variabile şi se refac, în noile condiții, calculele respective, până când se obține omogenitatea datelor înregistrate în şirul variabilei aleatoare X.

5.3. Prelucrarea primară a datelor statistice

Activitatea statistică este orientată în sensul cunoaşterii fenomenelor de masă din economie şi societate, răspunzând nevoii de informații la nivel micro şi macrosocial. Pentru a satisface aceste cerințe, cercetarea (investigația) statistică este organizată după un program riguros, cuprinzând mai multe etape, într-o succesiune logică. Astfel, **cercetarea statistică cuprinde totalitatea operațiilor de culegere, sistematizare, grupare, prelucrare, analiză şi interpretare a datelor şi informațiilor necesare pentru cunoaşterea fenomenelor şi proceselor economico-sociale**.

Acest proces amplu şi complex poate fi structurat în trei etape succesive: **observare, prelucrare şi analiză**.

Observarea este prima fază a demersului statistic şi are rolul de a asigura datele necesare investigației. Observarea presupune o analiză teoretică prealabilă, prin care este definită şi delimitată colectivitatea statistică supusă analizei, se identifică unitățile statistice şi sunt selectate caracteristicile de înregistrare care satisfac cel mai bine cerințele analizei.

În continuare este elaborat programul observării, care clarifică problemele metodologice şi organizatorice, apoi se trece la culegerea propriu-zisă a datelor.

Prelucrarea statistică începe cu centralizarea şi gruparea datelor observării, ca etape pregătitoare în vederea aplicării metodelor şi tehnicilor specifice statisticii. În urma prelucrării se obțin indicatorii statistici.

Mijloacele electronice de calcul permit în prezent prelucrarea unui mare volum de date individuale, care este concentrat în indicatori sintetici, mărimi ce evidențiază esența proceselor şi fenomenelor analizate. Datele statistice sunt prezentate sub formă de tabele, serii şi grafice.

Analiza şi interpretarea rezultatelor reprezintă ultima etapă a demersului statistic şi, totodată, încununarea eforturilor din etapele anterioare, de regulă, laborioase şi îndelungate. Comparând rezultatele prelucrării şi verificând ipotezele, se pot formula în final concluzii şi explicații asupra obiectului cercetării. Totodată, se fundamentează calculele de prognoză privind fenomenul analizat.

Material cu drept de autor

Deşi separate în timp şi spațiu, cele trei etape ale cercetării statistice sunt legate logic prin obiectul cercetării. Observarea trebuie să furnizeze un volum suficient de date de calitate corespunzătoare, care să asigure conținut real indicatorilor obținuți în etapa prelucrării. Alegerea celor mai potrivite metode de prelucrare şi utilizarea lor în funcție de particularitățile domeniului respectiv permit calcularea unor indicatori care surprind esența fenomenului investigat. Corectitudinea concluziilor finale se bazează pe autenticitatea datelor şi precizia prelucrării.

Datele statistice obținute în timpul observării sunt, de regulă, variate şi de volum mare. Ele trebuie sistematizate, centralizate şi grupate pentru a fi pregătite în vederea prelucrării. Numai în urma prelucrării statistice pot fi evidențiate trăsăturile şi tendințele esențiale din evoluția fenomenelor şi proceselor economico-sociale.

Prelucrarea primară a datelor statistice culese cuprinde operațiile de clasificare, grupare, centralizare, agregare, calcul de caracteristici derivate, precum şi construirea de tabele, serii şi grafice statistice.

Operațiile de prelucrare primară sunt permanent însoțite de **analiza datelor statistice**, în scopul identificării trăsăturilor definitorii ale datelor primare, al orientării procesului de prelucrare statistică, al alegerii celor mai potrivite metode şi tehnici statistice.

Prima fază a acestei analize este reprezentată de ordonarea, crescătoare sau descrescătoare, a valorilor distincte ale caracteristicii.

Aceste valori sunt însoțite, în cazul seriilor de frecvențe, de mărimi care arată numărul de apariții al fiecărei variante a caracteristicii.

Următoarea fază constă în clasificarea şi/sau gruparea datelor statistice, operații care permit ordonarea materialului statistic.

Considerând că ordonarea în serie crescătoare sau descrescătoare a datelor statistice nu ridică probleme, vom trece la operația de grupare a datelor statistice.

Gruparea datelor statistice constă în împărțirea unităților colectivității în ansambluri omogene, bine definite, după variația uneia sau mai multor caracteristici (numite factori de grupare). O grupă omogenă este formată din ansamblul unităților statistice care prezintă cel puțin o proprietate comună. Omogenitatea grupelor este asigurată şi de o variație redusă a valorilor individuale în cadrul fiecărei grupe.

Pentru a asigura o grupare corectă şi semnificativă se impune selecționarea caracteristicilor esențiale, cu caracter stabil. Alegerea caracteristicilor diferă în funcție de scopul analizei. De exemplu, dacă se analizează productivitatea muncitorilor dintr-o secție, vom urmări vechimea, calificarea, vârsta, dotarea tehnică etc., iar dacă se studiază potențialul unei unități economice, vom alege caracteristicile cifră de afaceri, profit, număr de personal, capital fix etc.

Există numeroase tipuri de grupări statistice, diferențiate după mai multe criterii.

1. În funcție de **numărul caracteristicilor** utilizate, grupările pot fi simple sau combinate.

Material cu drept de autor

• *Grupările simple* sunt cele realizate după o singură caracteristică de grupare. De exemplu: gruparea salariaților după venitul realizat sau gruparea județelor după numărul populației.

• *Grupările combinate* vizează două sau mai multe caracteristici de grupare, simultan. Gruparea se realizează etapizat: se alege o primă caracteristică după care se efectuează gruparea unităților colectivității şi fiecare grupă astfel obținută se împarte la rândul ei în subgrupe după variația celei de a doua caracteristici, apoi se repetă procedeul pentru cea de a treia caracteristică de grupare etc.

Întrucât o fărâmițare excesivă a colectivității ar anula semnificația grupării şi ar face imposibilă analiza, în practică se utilizează maximum trei caracteristici pentru gruparea combinată.

2. După **conținutul caracteristicii** de grupare deosebim grupări cronologice, teritoriale şi atributive.

• *Grupările cronologice* se referă la variația în timp a fenomenului analizat. În funcție de scopul analizei, mărimea intervalului de timp utilizat în grupare diferă: zi, lună, trimestru, an etc. De exemplu: producția zilnică realizată de muncitorii dintr-o secție, cheltuieli lunare de producție, profitul anual.

• *Grupările teritoriale* oglindesc variația în spațiu a unităților colectivității. De regulă, aceste grupări corespund unităților teritorial-administrative. Exemple: gruparea producției agricole pe județe, desfaceri cu amănuntul pe principalele oraşe ale țării, relații de import/export pe zone geografice şi țări.

• *Gruparea după o caracteristică atributivă* prezintă două variante:

- caracteristică exprimată prin cuvinte: gruparea producției pe ramuri economice, gruparea salariaților unei unități economice pe meserii etc.

- caracteristică exprimată numeric: gruparea societăților comerciale după cifra de afaceri, gruparea familiilor după numărul de copii etc.

3. După **modul de variație a caracteristicii**, gruparea se poate face pe variante, pe intervale egale de variație sau pe intervale inegale de variație.

• *Gruparea pe variante* presupune stabilirea unei grupe pentru fiecare valoare luată de caracteristica de grupare. Acest lucru este posibil atunci când numărul variantelor este redus. De exemplu: gruparea apartamentelor după numărul de camere, gruparea populației după starea civilă, gruparea salariaților după sex, gruparea studenților după tipul liceului absolvit etc.

• *Gruparea pe intervale de variație* se utilizează în cazul caracteristicilor numerice care înregistrează un număr mare de valori individuale diferite. Pentru a facilita analiza statistică, aceste valori sunt restrânse, sistematizate într-un număr redus de grupe. Fiecare grupă astfel constituită include unitățile colectivității pentru care valoarea caracteristicii se încadrează într-un anumit interval de valori.

Aceste intervale de variație pot fi de *mărimi egale* (de exemplu, gruparea după vârstă pe intervale de câte cinci ani, gruparea secțiilor după producție pe intervale de câte 10 unități monetare etc.) sau nu. Gruparea în

Material cu drept de autor

intervale de *mărimi inegale* este justificată numai atunci când repartiția valorilor individuale în cadrul colectivității este neuniformă sau atunci când o parte a colectivității statistice prezintă un interes deosebit, fiind necesară o analiză mai detaliată a acesteia. Gruparea pe intervale neegale permite evidențierea tipurilor calitative care se conturează în cadrul colectivității.

O importanță deosebită o are problema alegerii mărimii intervalului de grupare şi stabilirea numărului de grupe. *Alegerea numărului de grupe* trebuie să evite două erori frecvente:

- stabilirea unui *număr prea mare de grupe*, ceea ce conduce la fărâmițarea colectivității, cu consecințe negative atât pe planul identificării trăsăturilor esențiale ale fenomenului analizat, cât şi pe planul calculelor statistice ulterioare, care devin mai laborioase;

- stabilirea unui *număr prea mic de grupe* prezintă pericolul estompării deosebirilor calitative din cadrul structurii colectivității, alterând concluziile analizei.

Aşadar, alegerea numărului de grupe este o decizie dificilă care implică experiența şi talentul statisticianului. Acesta trebuie să aibă în vedere natura caracteristicii, amplitudinea variației valorilor înregistrate, scopul analizei statistice etc. Fiecare interval de grupare trebuie să cuprindă un număr suficient de mare de valori individuale.

Gruparea pe intervale egale presupune următoarele operațiuni:

- stabilirea caracteristicii de grupare;
- calcularea amplitudinii variației;
- stabilirea mărimii intervalului de grupare;
- precizarea limitelor superioare şi inferioare ale intervalelor de grupare;
- determinarea numărului unităților statistice care sunt incluse în fiecare interval.

Amplitudinea variației (A) se stabileşte ca diferență între valoarea maximă (x_{max}) şi valoarea minimă (x_{min}) înregistrată de caracteristica respectivă:

$$A = x_{max} - x_{min}$$

Mărimea intervalului de grupare (**h**) se determină pe baza raportului dintre amplitudinea variației (**A**) şi numărul de grupe (**h**) ales:

$$h = \frac{A}{k} = \frac{x_{max} - x_{min}}{k} \qquad (5.8)$$

Atunci când câtul împărțirii nu este un număr întreg se rotunjeşte în plus pentru a nu rămâne valori în afara ultimului interval de grupare.

Pentru determinarea mărimii intervalului de grupare se poate utiliza şi **formula lui Sturges,** recomandată în literatura de specialitate pentru colectivitățile de dimensiuni relativ mari, urmărind o distribuție apropiată de cea normală.

Astfel:

Material cu drept de autor

$$h = \frac{x_{\max} - x_{\min}}{1 + 3{,}322 . \lg n} \tag{5.9}$$

în care **n** - numărul unităților statistice din colectivitatea analizată.

Intervalele de grupare se definesc prin precizarea limitei inferioare şi superioare.

Determinarea primului interval de grupare porneşte de la valoarea minimă a caracteristicii (limita inferioară) la care se adaugă mărimea intervalului de grupare, obținându-se limita superioară.

În cazul caracteristicilor cu variație continuă, limita superioară a primului interval devine limita inferioară a celui de-al doilea interval.

Limita superioară a intervalului al doilea se obține adăugând mărimea intervalului la limita inferioară. Limita superioară a intervalului al doilea devine limita inferioară pentru al treilea interval şi procedeul continuă până când se precizează limitele tuturor celor **k** intervale.

Întrucât gruparea are *variație continuă*, este necesar să se precizeze care din cele două limite (inferioară sau superioară) este inclusă în interval. În acest fel ne asigurăm că o valoare situată la granița dintre două intervale va fi inclusă într-un singur interval, respectându-se unicitatea grupării.

În cazul caracteristicilor cu variație discretă, limita superioară a unui interval se diferențiază de limita inferioară a intervalului următor.

În practica statistică se întâlnesc deseori situații în care valorile extreme ale caracteristicii ($\mathbf{x_{min}}$ şi $\mathbf{x_{max}}$) sunt foarte îndepărtate de restul valorilor. În acest caz valorile extreme pot fi omise, iar primul şi ultimul interval devin intervale deschise (nu este precizată limita inferioară a primului interval şi nici limita superioară a ultimului interval). Atunci când prelucrările ulterioare o impun, aceste intervale pot fi închise, fiind considerate de lungime egală cu intervalele vecine. Lungimea fiecărui interval este dată de diferența dintre cele două limite.

Se recomandă ca limitele de interval să se exprime prin numere întregi.

Gruparea pe intervale neegale este preferabilă în cazul colectivităților de dimensiuni mari, cu structură neomogenă şi cu o amplitudine a variației foarte mare. O bună grupare pe intervale neegale depinde de acuratețea analizei calitative asupra structurii şi particularităților colectivității.

În cele ce urmează, vom ilustra modalitățile de grupare a datelor pe intervale egale şi neegale folosind următoarele mărimi ale producției realizate (număr bucăți) într-o lună de către 65 de muncitori ai unei unități industriale, dintr-o înregistrare inițială sub forma următoarelor date: **61, 66, 56, 71, 73, 67, 76, 69, 66, 77, 81, 69, 58, 50, 59, 74, 64, 71, 76, 63, 51, 73, 61, 64, 67, 64, 66, 69, 63, 72, 78, 82, 67, 57, 71, 83, 61, 73, 64, 58, 68, 62, 67, 63, 67, 69, 61, 79, 62, 68, 63, 67, 62, 68, 69, 66, 58, 72, 78, 67, 84, 66, 59, 73, 66**.

Gruparea pe intervale egale. Producția cea mai mică realizată în întreprindere a fost de 50 de bucăți, iar cea mai mare de 84 de bucăți. Amplitudinea variației (A) este:

Material cu drept de autor

A = x_{max} - x_{min} = 84 - 50 = 34 bucăți

Având în vedere amplitudinea variației şi volumul colectivității, am stabilit un număr de 7 grupe. Împărțind amplitudinea variației la numărul de grupe ales (**k**), determinăm mărimea intervalelor egale de variație (**h**):

$$h = \frac{x_{\max} - x_{\min}}{k} = \frac{A}{k} = \frac{34}{7} = 4{,}857 \cong 5\,bucati$$

Natura datelor ne indică să folosim intervale de variație discrete. Limitele intervalelor de grupare şi numărul muncitorilor din fiecare grupă (frecvențele absolute) sunt prezentate în tabelul 5.4.

Tabelul 5.4.

Gruparea muncitorilor după producția lunară realizată. Intervale egale:

Grupe de muncitori după producția realizată (bucăți)	Număr de muncitori	Producția cumulată (bucăți)
50 – 54	2	101
55 – 59	7	405
60 – 64	15	913
65 – 69	21	1407
70 – 74	10	754
75 – 79	6	462
80 – 84	4	328
Total	65	4370

Gruparea pe intervale neegale. În continuare se vor regrupa muncitorii astfel încât să se evidențieze trei tipuri calitative din punct de vedere al producției realizate: mic, mijlociu, mare. Folosind criteriul mediei, se defineşte tipul mijlociu prin reunirea a trei intervale: intervalul care conține media (65 - 69) şi cele două intervale alăturate (60 - 64 şi 70 - 74). Noua grupare este prezentată în tabelul 5.5.

Tabelul 5.5.

Gruparea muncitorilor după producția lunară realizată. Intervale neegale:

Producția lunară	Intervale (bucăți)	Număr muncitori	Producția cumulată (bucăți)
- mică	50 – 59	9	506
- mijlocie	60 – 64	46	3074
- mare	75 – 84	10	790
Total	-	65	4370

5.4. Serii de distribuție a frecvențelor

Observațiile înregistrate în prima fază a cercetării statistice sunt supuse unui proces de sistematizare, de ordonare.

Material cu drept de autor

În urma grupării după caracteristici atributive se obțin serii de distribuție (repartiție) a frecvențelor pe intervale de valori sau pe variante. Acestea oferă imaginea structurii colectivității, a repartizării unităților ei după intervalele de valori sau variantele caracteristicii de grupare.

Seria de distribuție de frecvențe este compusă din două şiruri de date:

• primul şir cuprinde variantele caracteristicii sau intervalele de valori;

• al doilea şir arată numărul unităților incluse în fiecare grupă astfel formată (frecvența, efectivul).

În funcție de natura caracteristicii de grupare, o serie de distribuție poate fi cantitativă sau calitativă, continuă sau discretă. Totodată, repartițiile pot fi teoretice, dacă reflectă o legitate matematică de repartiție a frecvențelor, sau empirice, dacă rezultă în urma prelucrării datelor reale.

Repartițiile empirice prezintă următoarele proprietăți:

► **omogenitatea** termenilor: variantele individuale sunt de aceeaşi natură şi cu valori apropriate, fiind determinate în cea mai mare măsură de acțiunea aceloraşi factori esențiali;

► **variabilitatea** valorilor individuale este dată de acțiunea mai puternică a unor factori întâmplători, care determină abaterea mărimilor individuale de la tendința centrală impusă de factorii esențiali;

► **independența** termenilor este efectul existenței distincte a unităților statistice în cadrul colectivității totale; fiecare unitate este rezultatul unei manifestări individualizate, diferite a fenomenului de masă;

► **forma** repartiției derivă din modalitatea specifică de combinare a influențelor factorilor esențiali şi neesențiali; există serii cu o repartiție relativ uniformă a frecvențelor şi altele cu unul sau mai multe puncte de concentrare. Aceste concentrări ale frecvențelor apar fie în jurul tendinței centrale, fie la unul sau ambele capete ale seriei.

O analiză completă a seriilor de distribuție a frecvențelor se bazează pe următorii indicatori:

• indicatori de **nivel** (nivel individual şi nivel totalizat sau valoare centralizată) şi de **frecvențe** (frecvențe absolute, frecvențe relative şi frecvențe cumulate);

• **indicatori medii:** media aritmetică, armonică, pătratică, geometrică;

• **indicatori medii de poziție**: mediană, cuartile, decile, modul;

• **indicatori simpli şi sintetici ai variației**: amplitudinea variației, abateri individuale, abatere medie liniară, abaterea medie pătratică (abaterea standard, abaterea tip), dispersia (varianța), coeficientul de variație;

• **indicatori ai asimetriei**;

• **indicatori ai concentrării**.

Acest sistem complex de indicatori este completat cu reprezentări grafice (histogramă, poligonul frecvențelor, poligonul frecvențelor cumulate) care pun în evidență forma repartiției şi orientează analiza.

Material cu drept de autor

Scopul analizei şi particularităţile seriei de repartiţie studiate determină indicatorii cei mai potriviţi pentru fiecare caz în parte.

În cazul datelor negrupate, indicatorii de nivel sunt chiar valorile individuale ale unei caracteristici ($\mathbf{x_i}$ unde **i** ia valori de la **1** la **n**). În cazul seriilor de distribuţie unidimensionale, ca indicatori de nivel ($\mathbf{x_i}$ unde **i** ia valori de la **1** la **k**, în cazul unei distribuţii de frecvenţe cu **k** grupe) se utilizează variantele în cazul grupării pe variante şi centrele de intervale (calculate ca medie aritmetică simplă a limitelor fiecărui interval) în cazul grupării pe intervale. Nivelul totalizat al caracteristicii se calculează în mod diferit în funcţie de tipul seriei.

Pentru o serie de date negrupate, nivelul totalizat se obţine prin însumarea tuturor valorilor individuale:

$$\sum_{i=1}^{n} x_i$$

Pentru date grupate se poate calcula nivelul totalizat al fiecărei grupe şi pe total în funcţie de datele disponibile.

Dacă se cunosc valorile individuale din care s-a efectuat gruparea:

- nivelul totalizat al grupei:

$\sum_{j=1}^{n_i} x_{ij}$ unde $\mathbf{n_i}$ este volumul grupei iar $\mathbf{x_{ij}}$ valorile individuale din cadrul grupei respective;

- nivelul centralizat pe total:

$\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}$ unde **k** este numărul de grupe.

În cazul în care nu se cunosc valorile individuale din cadrul fiecărei grupe:

- nivelul totalizat al grupei: $\mathbf{x_i n_i}$, unde $\mathbf{n_i}$ este volumul grupei iar $\mathbf{x_i}$ indicatorul de nivel al grupei;

- nivelul centralizat pe total:

$\sum_{i=1}^{k} x_i n_i$ unde **k** este numărul de grupe.

Indicatorii de nivel totalizat se pot reprezenta grafic prin diagrame de volum al caracteristicii (pătrat, cerc, dreptunghi) a căror suprafaţă este proporţională cu valoarea de reprezentat grafic sau prin diagrama prin coloane nelipite.

Material cu drept de autor

Tabelul 5.6.

Gruparea muncitorilor unei firme după nivelul dotării tehnice (date convenționale)

Grupe de muncitori după dotarea tehnică (u.m. capital fix pe muncitor)	Număr de muncitori (n_i)	Frecvențe absolute cumulate	
		Crescător F_i^A	Descrescător F_i^D
0	1	2	3
1,5 – 2,0	14	14	150
2,0 – 2,5	18	32	136
2,5 – 3,0	30	62	118
3,0 – 3,5	48	110	88
3,5 – 4,0	25	135	40
4,0 – 4,5	15	150	15
Total	150	-	-

Pentru seriile de distribuție (repartiție) se pot calcula următorii indicatori de frecvențe: **frecvențe absolute**; **frecvențe relative**; **frecvențe cumulate**.

Frecvențele absolute (n_i) rezultă din operația de grupare a unităților colectivității. Numărul unităților statistice care aparțin unei grupe (clase) reprezintă frecvența grupei respective.

Întrucât frecvențele absolute se exprimă în unități concrete de măsură (număr de muncitori, de firme, de persoane, de magazine etc.) nu se pot compara între ele serii statistice.

Reprezentarea grafică a frecvențelor absolute se face cu ajutorul histogramei şi a poligonului frecvențelor.

Se va exemplifica în continuare modul de construire a histogramei şi a poligonului frecvențelor pentru intervale de grupare egale, pe baza datelor din tabelul 5.6.

Histograma pentru intervale egale se obține construind un număr de coloane egal cu numărul grupelor, având baze egale pe Ox (variația este continuă) şi înălțimea dată de frecvențele grupelor pe care le reprezintă, figura 5.2.

Exemplul de calcul 5.2. Pentru o societate comercială (SC) se cunosc următoarele date statistice privind vechimea în muncă a angajaților, tabelul 5.2:

Tabelul 5.2. Date inițiale.

Grupe de salariați după vechime, ani	<5	5-10	10-15	15-20	20-25	25-30	30-35	≥35
Număr salariați	10	40	60	80	50	30	20	10

Material cu drept de autor

Se cere, reprezentarea grafică a seriei folosind frecvențele absolute şi cele cumulate.

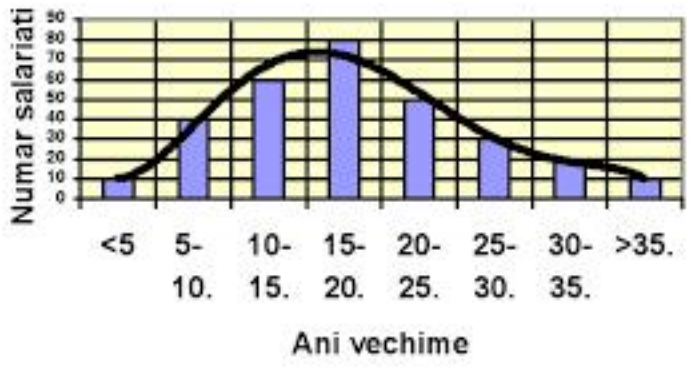


Fig.5.2. Histograma şi curba de distribuție pentru seria statistică din tabelul 5.2.

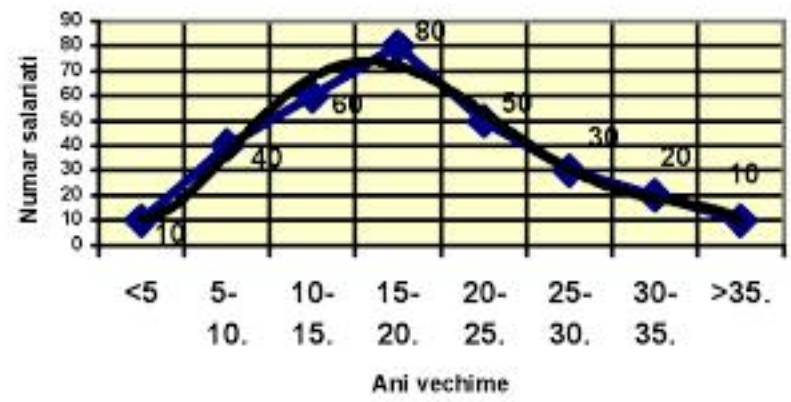


Fig.5.3. Poligonul frecvențelor absolute şi curba de distribuție pentru seria statistică din tabelul 5.2.

Aceste tipuri de grafice pun în evidență forma de variație a caracteristicii statistice şi gradul de asimetrie a seriei.

Poligonul frecvențelor absolute cumulate, ascendent şi descendent, corespunzător gruprii pe intervale egale este prezentat în figura 5.4.

Frecvențele relative arată importanța relativă a fiecărei grupe, punând în evidență structura colectivității. Folosirea acestui indicator permite să se compare seriile empirice între ele sau cu distribuțiile teoretice.

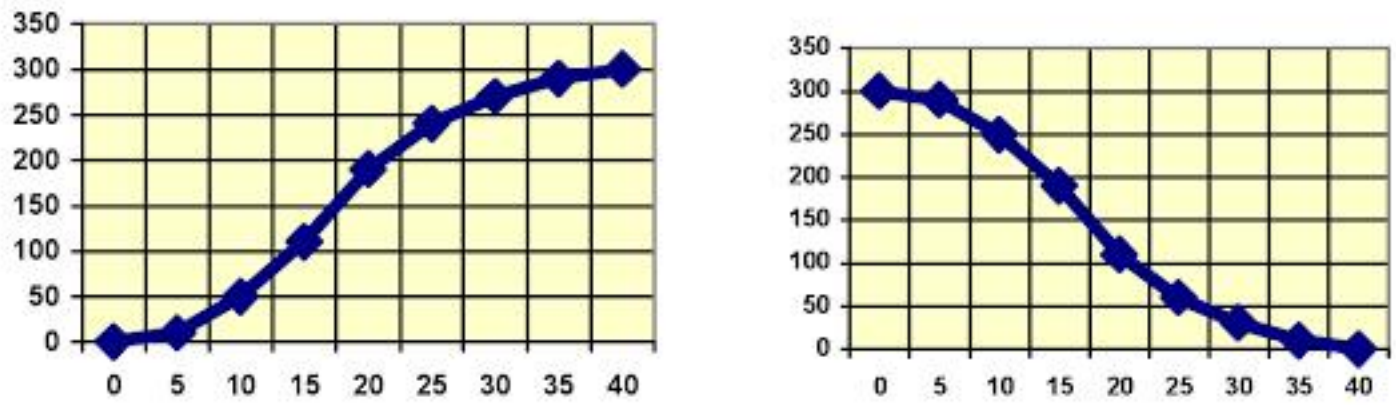


Fig.5.4. Repartiția frecvențelor absolute cumulate, ascendent şi descendent

5.5. Indicatori ai tendinței de grupare a datelor seriilor statistice

5.5.1. Indicatori de medii: media aritmetică, armonică, pătratică, geometrică

Statistica încearcă să redea ceea ce este tipic, comun şi general în evoluția fenomenelor şi proceselor economice. Variabilitatea deosebită a acestora în formele lor individuale de manifestare impune găsirea unor

Material cu drept de autor

indicatori sintetici care să reunească valorile individuale, exprimând printr-o măsură unică esența fenomenului.

Indicatorii medii răspund acestui deziderat, sintetizând ceea ce este comun, tipic în manifestarea fenomenului considerat. Ei măsoară influența cauzelor esențiale, înlăturând variațiile întâmplătoare în evoluția fenomenelor şi proceselor.

Media aritmetică. Valoarea medie aritmetică **M(X)** a unei variabile aleatoare **X**, care poate lua un număr finit de valori:

$$X:\begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

este prin definiție:

$$M(X) = p_1x_1 + p_2x_2 + \dots + p_nx_n \qquad (5.10)$$

Valoarea medie se mai notează **μ, m, $\bar{x}$** şi poate purta numele de ***speranță matematică***.

Dacă se au în vedere mai multe variabile aleatoare împreună se obține un sistem de variabile aleatoare.

Se consideră două variabile aleatoare: **X** şi **Y**:

$$X:\begin{pmatrix} x_1 & x_2 & \dots & x_m \\ p_1 & p_2 & \dots & p_m \end{pmatrix} \text{ şi } Y:\begin{pmatrix} y_1 & y_2 & \dots & y_n \\ q_1 & q_2 & \dots & q_n \end{pmatrix}$$

Perechea **(X, Y)** are următoarea distribuție:

$$X,Y:\begin{pmatrix} (x_1y_1) & (x_1y_2) & \dots & (x_1y_n) & (x_2y_1) & (x_2y_2) & \dots & (x_my_n) \\ p_{11} & p_{12} & \dots & p_{1n} & p_{21} & p_{22} & \dots & p_{mn} \end{pmatrix} \qquad (5.11)$$

unde p_{ij} reprezintă probabilitatea de obținere a perechii (x_i, y_j).

Este evident că dacă cele două variabile aleatoare sunt independente, atunci, în baza regulii de înmulțirea probabilităților, va rezulta:

$$p_{ij} = p_iq_j \qquad (5.12)$$

Sistemului de variabile aleatoare i se pot aplica diferite operații ca de exemplu, înmulțirea cu o constantă **K**, adunarea, înmulțirea etc.

Astfel, dacă se consideră expresia **X + Y** – suma celor două variabile aleatoare considerate anterior – va rezulta tot o variabilă aleatoare care poate lua diferite valori ca: x_1+y_1, x_1+y_2, x_2+y_1, x_2+y_2,... deci, în general, valori x_i+y_j, cu probabilitățile respective p_{11}, p_{12}, p_{21}, p_{22},...,p_{ij}.

Este clar că problema se pune de a calcula valori medii şi pentru ***sisteme de variabile aleatoare***.

Valoarea medie a sumei de două variabile. Notând cu **i = 1**, 2,...h, indicii valorilor x_i, luate de **X** şi cu **j = 1, 2,...,k** indicii valorilor y_j, luate de **Y**, valoarea medie pentru suma **X + Y**, conform definiției din relația (5.10) va fi:

Material cu drept de autor

$$M(X+Y) = p_{11}(x_1 + y_1) + p_{12}(x_1 + y_2) + \ldots +$$

$$+ p_{h1}(x_h + y_1) + p_{h2}(x_h + y_2) + \ldots + p_{hk}(x_h + y_k)$$

Termenii din partea doua a expresiei de mai sus pot fi regrupați şi se va obține:

$$M(X+Y) = x_1(p_{11} + p_{12} + \ldots + p_{1k}) + \ldots +$$

$$+ x_h(p_{h1} + p_{h2} + \ldots + p_{hk}) + \ldots + y_k(p_{1k} + \ldots + p_{hk})$$

de unde, ținând seama de relația (5.11), rezultă:

$$M(X+Y) = p_1x_1 + \ldots + p_hx_h + q_1y_1 + \ldots + q_ky_k$$

şi cum în baza relației (5.10):

$$p_1x_1 + \ldots + p_hx_h = M(X) \quad \text{şi} \quad q_1y_1 + \ldots + q_ky_k = M(Y)$$

va rezulta:

$$M(X+Y) = M(X) + M(Y) \qquad (5.13)$$

Se poate, deci, formula următoarea teoremă:

Valoarea medie a unei sume de două variabile dintr-un sistem de variabile aleatoare este egală cu suma valorilor medii ale celor două variabile aleatoare.

Prin generalizare, aplicându-se acelaşi raționament ca în cazul a două variabile aleatoare – media sumei sistemului de variabile independente este:

$$M(X_1 + X_2 + \ldots + X_n) = M(X_1) + M(X_2) + \ldots + M(X_n) \qquad (5.14)$$

Prin extensie, se poate formula, pentru cazul considerat, următoarea teoremă: ***Valoarea medie a sumei unui număr finit de variabile aleatoare este egală cu suma valorilor medii ale fiecărei variabile***.

Valoarea medie a unei variabile din sistem înmulțită cu o constantă **K** va fi:

$$M(KX) = KM(X) \qquad (5.15)$$

Se poate demonstra uşor că valoarea medie a produsului a două variabile aleatoare independente $\mathbf{X_1}$ şi $\mathbf{X_2}$ este tot o variabilă aleatoare **Z** determinată cu relația:

$$\mathbf{Z = X.Y} \qquad (5.16)$$

Dacă extindem raționamentul pentru **n** variabile aleatoare independente $\mathbf{X_1, X_2, \ldots, X_n}$, atunci se obține relația:

$$M(X_1X_2 \ldots X_n) = M(X_1)M(X_2) \ldots M(X_n) \qquad (5.17)$$

Prin urmare se poate enunța următoarea teoremă: ***Valoarea medie a unui produs de variabile aleatoare independente este egală cu produsul valorilor medii ale variabilelor aleatoare considerate***.

Material cu drept de autor

Dacă **X** este o variabilă aleatoare de tip continuu, având densitatea de repartiție **f**, atunci numărul:

$$M(X) = \int_{-\infty}^{+\infty} x.f(x).dx \qquad (5.18)$$

se numeşte valoare medie a variabilei aleatoare **X** dacă integrala din membrul drept este convergentă.

Dacă variabila aleatoare **X** ia valori numai în intervalul **(a, b)**, **($-\infty$ < a < b <$+\infty$)** atunci **f(x) = 0** pentru **x$\notin$(a, b)** şi deci:

$$M(X) = \int_{a}^{b} x.f(x).dx \qquad (5.19)$$

Valoarea medie este una dintre cele mai importante caracteristici numerice ataşate variabilei aleatoare. Rolul acestor caracteristici numerice este de a ne permite, în anumite situații, să tragem unele concluzii asupra variabilelor aleatoare fără a apela la legile lor de probabilitate – de cele mai multe ori foarte dificil sau imposibil de obținut. De multe ori, cunoaşterea tipului de interdependență dintre variabilele unui şir şi a unora din caracteristicile lor numerice ne permit să tragem concluzii importante din punct de vedere practic sau teoretic.

După tehnica de calcul, media aritmetică este de două feluri şi anume: simplă şi ponderată.

Media aritmetică simplă. Fiind dată seria statistică $x_1, x_2, \ldots, x_n$, media aritmetică simplă se determină cu expresia:

$$m_a = \frac{x_1 + x_2 + \ldots + x_n}{n} \qquad (5.20)$$

Media aritmetică prezintă următoarele proprietăți:

1. $\min\limits_{1 \le i \le n} x_i \le m_a \le \max\limits_{1 \le i \le n} x_i$
2. $x_1 + x_2 + \ldots + x_n = \underbrace{m_a + m_a + \ldots + m_a}_{de\ n\ ori} = n.m_a$
3. $\sum\limits_{i=1}^{n} (x_i - m_a) = 0$
4. $\sum\limits_{i=1}^{n} (x_i - v) = n.(m_a - v)$, oricare ar fi v.

Material cu drept de autor

5. $\sum_{i=1}^{n}(x_i - v)^2 \geq \sum_{i=1}^{n}(x_i - m_a)^2$, oricare ar fi v.

Exemplu de calcul. Într-un atelier, în timp de 8 ore, 8 muncitori realizează următorul număr de piese:

Muncitorul	A	B	C	D	E	F	G	H
Nr. de piese	20	22	30	25	32	18	21	32

Se cere să se determine numărul mediu de piese produse de un muncitor. Se va determina cu expresia:

$$m_a = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{20+22+30+25+32+18+21+32}{8} = \frac{200}{8} = 25$$

Media aritmetică ponderată. Dacă se notează nivelele individuale cu $\mathbf{x_1, x_2,\ldots,x_p}$, iar frecvențele absolute cu $\mathbf{n_1, n_2,\ldots,n_p}$, formula mediei aritmetice ponderate va fi:

$$m_a = \frac{x_1 n_1 + x_2 n_2 + \ldots + x_p n_p}{n_1 + n_2 + \ldots + n_p} = \frac{\sum_{i=1}^{p} x_i n_i}{\sum_{i=1}^{p} n_i} \qquad (5.21)$$

Exemplu de calcul 5.3. Considerând seria statistică $\mathbf{x_i}$, $\mathbf{n_i}$ a distribuției notelor la o lucrare de statistică:

$\mathbf{x_i}$	3	4	5	6	7	8	9	10
$\mathbf{n_i}$	1	3	4	7	5	4	4	2

Se cere să se determine valoarea mediei obținută la lucrarea de statistică:

Aplicând relația de mai sus rezultă:

$$m_a = \frac{x_1 n_1 + x_2 n_2 + \ldots + x_p n_p}{n_1 + n_2 + \ldots + n_p} = \frac{\sum_{i=1}^{p} x_i n_i}{\sum_{i=1}^{p} n_i} =$$

$$= \frac{1.3+4.3+5.4+6.7+7.5+8.4+9.4+10.2}{1+3+4+7+5+4+4+2} = \frac{210}{30} = 7$$

Metode simplificate de calcul pentru media aritmetică.

1. Media calculată din variantele caracteristicii micşorate cu o constantă „a" este mai mică decât media reală cu constanta „a"; simbolic această proprietate este sintetizată de inegalitatea:

Material cu drept de autor

$$\frac{\sum_{i=1}^{n}(x_i - a)f_i}{\sum_{i=1}^{n} f_i} < \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} \quad cu \quad "a"$$

în care: partea stângă a inegalității reprezintă media obținută din variantele micşorate cu constanta „a"; partea dreaptă – media reală.

Pentru a le egala sunt două posibilități:

- să se mărească partea stângă cu „a";
- să se micşoreze partea dreaptă cu „a".

Se alege prima posibilitate:

$$\frac{\sum_{i=1}^{n}(x_i - a)f_i}{\sum_{i=1}^{n} f_i} + a = \bar{x} \qquad (5.22)$$

care se verifică, întrucât dezvoltând expresia din stânga relației, se ajunge la:

$$\frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} - \frac{a\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} f_i} + a = \bar{x}$$

Constanta „a" se alege dintre valorile $x_1, x_2, \ldots, x_n$, fiind, de regulă, variabila cu frecvența cea mai mare.

2. Media calculată din variantele caracteristicii micşorate prin împărțire la o constantă „k" este mai mică decât media reală de „k" ori.

$$\frac{\sum_{i=1}^{n}\left(\frac{x_i}{k}\right)f_i}{\sum_{i=1}^{n} f_i} < \frac{a\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} \quad de\ "k"\ ori$$

k reprezintă mărimea intervalului de grupare.

Pentru ale egala se alege una din posibilități şi anume:

$$\frac{\sum_{i=1}^{n}\left(\frac{x_i}{k}\right)f_i}{\sum_{i=1}^{n} f_i} k = \bar{x}$$

Material cu drept de autor

Dezvoltând membrul stâng al relației se ajunge la:

$$\frac{\frac{1}{k}\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} k = \bar{x} \qquad (5.23)$$

3. Metoda combinată, a celor două de mai sus:

$$\frac{\sum_{i=1}^{n}\left(\frac{x_i - a}{k}\right) f_i}{\sum_{i=1}^{n} f_i} k + a = \bar{x} \qquad (5.24)$$

Întrucât constantele „a" şi „k" sunt elemente ale seriilor de distribuție, calculul simplificat se aplică numai mediei aritmetice ponderate, caz în care seriile de distribuție trebuie să fie construite după intervale egale de variație.

De multe ori, în practica statistică, se întâlnesc serii de distribuții la care frecvențele absolute f_1, f_2,...,f_n au niveluri relativ mari, afectând întrucâtva calculul mediei. Pentru a facilita şi mai mult simplificarea calculului se poate apela la încă o proprietate a mediei aritmetice, care vizează frecvențele şi nu variabilele caracteristicii şi anume:

- dacă se micşorează frecvențele f_1, f_2,...,f_n prin împărțire la o constantă „c" aleasă arbitrar, nivelul mediei aritmetice nu se schimbă, adică:

$$\frac{\sum_{i=1}^{n} x_i \frac{f_i}{c}}{\sum_{i=1}^{n} \frac{f_i}{c}} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i}$$

întrucât:

$$\frac{\frac{1}{c}\sum_{i=1}^{n} x_i f_i}{\frac{1}{c}\sum_{i=1}^{n} f_i} = \bar{x}$$

În condițiile combinării celor trei proprietăți, relația de calcul simplificat al mediei aritmetice este:

Material cu drept de autor

$$\frac{\sum_{i=1}^{n}\left(\frac{x_i - a}{k}\right)\frac{f_i}{c}}{\sum_{i=1}^{n}\frac{f_i}{c}}k + a = \bar{x} \qquad (5.25)$$

Se va considera următorul exemplu:

Exemplu de calcul 5.4. Se cere să se determine media aritmetică a salariului mediu pentru un grup de salariați care realizează veniturile lunare menționate în coloanele 1 şi 2 din tabelul următor:

Salariul lunar u.m.	Număr salariați	Centrul intervalelor	$x_i f_i$	$\frac{x_i - a}{k}$	$\left(\frac{x_i - a}{k}\right) f_i$
1	2	3	4	5	6
1500-1700	4	1600	6400	-2	-8
1700-1900	14	1800	25200	-1	-14
1900-2100	20	**2000**	40000	0	0
2100-2300	10	2200	22000	1	10
2300-2500	7	2400	16800	2	14
2500-2700	4	2600	10400	3	12
Peste 2700	1	2800	2800	4	4
Total	60	x	123600	x	18
Observație. Limita superioară este inclusă în interval.(a=2000; k=200)					

Rezolvare. Se observă că distribuția salariaților este construită după intervale de variație. Pentru toate cazurile de acest gen, în calculul mediei aritmetice ponderate se folosesc ca variabile $x_1, x_2, \ldots, x_n$ ***centrele intervalelor de variație***. Ele se determină ca medii aritmetice simple din cele două limite ale fiecărui interval. De exemplu:

$$x_1 = \frac{1500 + 1700}{2} = 1600 \text{ etc.}$$

Pentru intervalele deschise la unul din capete, centrul se determină astfel:

- dacă intervalul deschis este la capătul inferior (cazul distribuțiilor deschise la începutul lor), se determină mai întâi limita inferioară a intervalului respectiv, scăzând din limita superioară mărimea intervalului de grupare (**k**) şi apoi centru intervalului respectiv ca semisumă a celor două limite componente;

- dacă intervalul este deschis la capătul superior (cazul distribuțiilor deschise la sfârşitul lor), se determină mai întâi limita superioară a acestui interval prin adăugarea la limita inferioară a lui **k** şi apoi centrul său prin semisuma acestor limite.

Practic, pentru determinarea centrului intervalelor deschise există mai multe posibilități.

Material cu drept de autor

Aplicând relația de calcul obişnuit salariul mediu se determină astfel:

$$\bar{x} = \frac{1600.4 + 1800.14 + ... + 2800.1}{60} = 2060\ u.m.$$

Interpretare. În medie, unui salariat i-a revenit un salariu lunar de 2060 u.m.

Aplicând relația de calcul simplificat se va obține acelaşi rezultat, însă mult mai operativ, deşi formula pare la prima vedere mai complicată:

$$\bar{x} = \frac{18}{60}200 + 2000 = 2060\ u.m.$$

Operativitatea calculului simplificat este dată de faptul că odată stabilită poziția lui „a" în cadrul seriei, nivelurile rapoartelor $\frac{x_i - a}{k}$ se trec automat astfel: zero în dreptul lui **„a"**, **-1, -2, -3,...** deasupra şi **1, 2, 3,...** sub zero.

Se va verifica acest lucru prin exemplul prezentat în tabelul precedent. De exemplu:

$$\frac{x_1 - a}{k} = \frac{1600 - 2200}{200} = -2$$

.

.

$$\frac{x_7 - a}{k} = \frac{2800 - 2000}{200} = 4$$

Aşadar, avantajul pe care-l are utilizarea procedeului simplificat este dublu:

- se micşorează considerabil nivelul indicatorilor folosiți în calculul mediei, ceea ce facilitează efectuarea operațiunilor aferente;

- se pot cunoaşte direct, fără calcule, valorile rapoartelor $\frac{x_i - a}{k}$ pentru orice tip de distribuție, indiferent de locul pe care îl ocupă constanta „a" în cadrul seriei respective.

Datorită acestor însuşiri, procedeul de calcul simplificat al mediei aritmetice ponderate este preferat procedeul obişnuit.

Mediile: armonică, geometrică, pătratică, cronologică. Asupra variabilelor aleatoare se pot face anumite operații, obținându-se alte variabile aleatoare care au distribuțiile lor proprii cu valorile medii corespunzătoare. O variabilă aleatoare poate fi ridicată la o putere de ordinul **k**, în care caz se pune problema de a-i calcula valoarea medie.

Prin definiție, expresia:

Material cu drept de autor

$$M(X^k) = p_1 x_1^k + p_2 x_2^k + \ldots + p_n x_n^k \qquad (5.26)$$

este tot o valoare medie a variabilei **X** însă ridicată la puterea **k**.

Unei asemenea valori medii exprimată cu relația (5.26) i se spune ***moment de ordinul k al variabilei aleatoare X***.

Dacă se extrage radicalul de ordinul **k** din expresia (5.26), se obține altă medie de variabilă aleatoare:

$$M_k = \sqrt[k]{M(X^k)} = \sqrt[k]{p_1 x_1^k + p_2 x_2^k + \ldots + p_n x_n^k}$$

care se numeşte ***valoare medie de ordinul k a aceleiaşi variabile aleatoare***.

Se poate observa simplu că ***speranța matematică*** **M(X)** nu este decât un caz particular al expresiei (5.26), pentru **k = 1**; este vorba de momentul de ordinul unu al variabilei aleatoare **X**.

Media armonică. Pentru **k = -1**, avem:

$$M_{-1} = \left(\frac{p_1}{x_1} + \frac{p_2}{x_2} + \ldots + \frac{p_n}{x_n} \right)^{-1}$$

sau:

$$m_h = \frac{p_1 + p_2 + \ldots + p_n}{\frac{p_1}{x_1} + \frac{p_2}{x_2} + \ldots + \frac{p_n}{x_n}}$$

numită ***medie armonică*** a valorilor x_i.

Sub altă formă demonstrativă, ***media armonică*** se poate obține în felul următor. Se pleacă de la definiția acesteia că media armonică reprezintă media mărimilor inverse a termenilor seriei de valori statistice. Considerând seria statistică $x_1, x_2, \ldots, x_n$, media armonică se determină cu expresia:

$$m_h = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}} = \frac{n}{\sum \frac{1}{x_i}}, \quad i = \overline{1,k}$$

Dacă $x_1, x_2, \ldots, x_n$ apar de $p_1, p_2, \ldots, p_n$ ori, atunci:

$$m_h = \frac{p_1 + p_2 + \ldots + p_n}{\frac{p_1}{x_1} + \frac{p_2}{x_2} + \ldots + \frac{p_n}{x_n}}$$

Media armonică se utilizează pentru calculul salariului mediu pe o unitate economică când se cunosc salariile medii şi fondul de salarii pe fiecare subunitate economică sau calculul recoltei medii la hectar când se cunosc recoltele la hectar şi totalul recoltei pe fiecare fermă sau prețul mediu al unui produs când se cunoaşte prețul pe anumite perioade.

Material cu drept de autor

Media armonică poate fi: simplă sau ponderată. Media armonică simplă se foloseşte când:

$$x_1 f_1 = x_2 f_2 = \ldots = x_n f_n$$

adică atunci când nivelul variabilei aleatoare se menține constant la fiecare grupă.

Metodologie de calcul. Ținând seama de cerințele impuse de condiția matematică de semnificație şi considerând $\mathbf{m_h}$ – simbolul mediei armonice, filiera calculului mediei armonice simple este identică cu cea de la media aritmetică simplă, cu singura deosebire că se foloseşte în calcul inversul variabilelor caracteristicii:

Media armonică simplă se determină cu relația:

$$\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n} = \sum_{i=1}^{n} \frac{1}{x_i} = \frac{n}{m_h} \qquad (5.27)$$

de unde rezultă:

$$m_h = \frac{1}{\dfrac{\dfrac{1}{x_1} + \dfrac{1}{x_2} + \ldots + \dfrac{1}{x_n}}{n}} = \frac{n}{\dfrac{1}{x_1} + \dfrac{1}{x_2} + \ldots + \dfrac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \dfrac{1}{x_i}}$$

Media armonică ponderată se utilizează în situația în care:

$$x_1 f_1 \neq x_2 f_2 \neq \ldots \neq x_n f_n$$

şi se determină cu relația:

$$m_h = \frac{1}{\dfrac{\dfrac{1}{x_1} x_1 f_1 + \dfrac{1}{x_2} x_2 f_2 + \ldots + \dfrac{1}{x_n} x_n f_n}{x_1 f_1 + x_2 f_2 + \ldots + x_n f_n}} =$$

$$= \frac{x_1 f_1 + x_2 f_2 + \ldots + x_n f_n}{\dfrac{1}{x_1} x_1 f_1 + \dfrac{1}{x_2} x_2 f_2 + \ldots + \dfrac{1}{x_n} x_n f_n} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} \dfrac{1}{x_i} x_i f_i} \qquad (5.28)$$

Exemplu de calcul 5.5. Se cere să se determine recolta medie la hectar pe suprafața unei ferme agricole cunoscând c recolta medie pe cele trei lanuri cultivate şi recoltele totale următoare:

Lanul	L_1	L_2	L_3	
Recolta medie	2,5 t	3 t	4 t	
Recolta totală	25 t	24 t	28 t	77 t

Material cu drept de autor

Rezolvare. Recolta medie pe fermă va fi:

$$m_h = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} \frac{1}{x_i} x_i f_i} = m_h = \frac{25+24+28}{\frac{1}{2,5}25 + \frac{1}{3}24 + \frac{1}{4}28} = \frac{77}{25} = 3,08t$$

Se observă că dacă se calcula media aritmetică ponderată a recoltelor medii se obținea:

$$\frac{2,5.25+3.24+4.28}{77} = 3,3t$$

fapt ce era fals, întrucât dacă se calculează direct raportul dintre totalul producției (77 t) şi numărul total de hectare rezultă (25:2,5 + 24:3 + 28:4 =25 ha) 3,08 t.

Media geometrică. Când **k = 0**, momentul M_0 se prezintă sub formă nedeterminată. Însă, aplicând logaritmii şi regula lui L'Hôpital, se demonstrează uşor că:

$$M_0 = \sqrt[p]{x_1^{p_1}.x_2^{p_2}...x_n^{p_n}}\ ;\ \ p = \sum_{i=1}^{n} p_i \qquad (5.29)$$

numită ***medie geometrică***.

Dacă avem seria numerelor reale pozitive nenule $x_1, x_2,...,x_n$, atunci ***media geometrică*** va fi:

$$m_g = \sqrt[n]{x_1.x_2....x_n}$$

Media geometrică prezintă următoarele proprietăți:

1. $\min\limits_{1\le i\le n} x_i \le m_g \le \max\limits_{1\le i\le n} x_i$
2. $x_1.x_2...x_n = \underbrace{m_g.m_g...m_g}_{de\ n\ ori}$
3. $\begin{cases} \frac{x_1}{m_g} \times \frac{x_2}{m_g} \times ... \times \frac{x_k}{m_g} = \frac{m_g}{x_{k+1}} \times \frac{m_g}{x_{k+2}} \times ... \times \frac{m_g}{x_n} \\ unde\ x_i < m_g,\ pentru: i = \overline{1,k}\ şi\ m_g < x_{k+j} \\ pentru\ \ j = \overline{1,n-k} \end{cases}$

Media geometrică numită şi **medie logaritmică** se calculează pe baza mărimilor relative ale dinamicii în cadrul seriilor cronologice. Cu ajutorul ei se determină ritmurile medii de creştere a populației, producției, circulației

Material cu drept de autor

mărfurilor, prețurilor, venitului național etc. Media geometrică este de două feluri: simplă şi ponderată.

Media geometrică simplă se determină pe baza expresiei:

$$m_g = \sqrt[n]{x_1.x_2...x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i} \qquad (5.30)$$

Exemplu de calcul 5.6. Se cere să se determine coeficientul mediu de creştere lunară a volumului desfacerilor de mărfuri a unui supermagazin pe baza următoarelor date:

Lunile	ianuarie	februarie	martie	aprilie	mai	iunie
Volumul desfacerilor (u.m)	300	320	350	400	420	450
Coeficientul de creştere cu baza în lanț	100	1,067	1,094	1,143	1,050	1,071

Rezolvare. Se determină, la început, coeficienții de creştere cu bază în lanț:

februarie: $320/300 = 1{,}067$; martie: $350/320 = 1{,}094$ etc.

Se aplică formula mediei geometrice:

$$m_g = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[5]{1{,}067 \times 1{,}094 \times 1{,}143 \times 1{,}050 \times 1{,}071} = \sqrt[5]{1{,}5004}$$

Aplicându-se calculul logaritmic obținem:

$$\log m_g = \frac{\log 1{,}5004}{5} = \frac{0{,}17609}{5} = 0{,}035218$$

Calculând antilogaritmul, rezultă:

$m_g = 1{,}084 \times 100 = 108{,}4\%$, adică ritmul mediei lunare de creştere a volumului desfacerilor de mărfuri a fost de 8,4%.

Media geometrică ponderată. Se determină cu expresia:

$$m_g = \sqrt[\sum_{i=1}^{k} n_i]{\prod_{i=1}^{k} x_i^{n_i}} \qquad (5.31)$$

Exemplu de calcul 5.7. Să se determine nivelul mediu de creştere a productivității muncii într-o unitate economică, considerându-se creşterile parțiale conform datelor din tabelul următor:

Perioada	1-5	6-10	11	12	13	14	15	16
Indicii dinamici	Indicii medii anuali 1,138	Indicii medii anuali 1,112	1,12	1,12	1,15	1,15	1,12	1,11

Rezolvare. Pentru a determina nivelul mediu al indicilor de creştere pe perioada de 16 ani, întrucât datele din tabel se referă la intervale de timp diferite, se va folosi relația de calcul a mediei geometrice ponderate:

Material cu drept de autor

$$m_g = \sqrt[\sum_{i=1}^{k} n_i]{\prod_{i=1}^{k} x_i^{n_i}} = \sqrt[16]{1{,}138^{5}.1{,}112^{5}.1{,}12.1{,}12.1{,}15.1{,}15.1{,}12.1{,}11}$$

$$\log m_g = \frac{5\log 1{,}138 + 5\log 1{,}112 + \log 1{,}12 + \ldots + \log 1{,}11}{16} = 0{,}05159$$

$$m_g = 1{,}126$$

Din calcule rezultă că în fiecare an al perioadei de 16 ani, productivitatea muncii creşte în medie, de 1,126 ori, față de anul precedent.

Media pătratică. Momentul de ordinul al doilea al unei variabile aleatoare este prin definiție:

$$M(X^2) = p_1 x_1^2 + p_2 x_2^2 + \ldots + p_n x_n^2 \qquad (5.32)$$

Expresia:

$$M_2 = \sqrt{M(X^2)} = \sqrt{p_1 x_1^2 + p_2 x_2^2 + \ldots + p_n x_n^2} \qquad (5.33)$$

sau:

$$m_p = \sqrt{\frac{p_1 x_1^2 + p_2 x_2^2 + \ldots + p_n x_n^2}{p_1 + p_2 + \ldots + p_n}} \qquad (5.34)$$

se numeşte ***valoare medie pătratică***.

Media pătratică este de două feluri: simplă şi ponderată.

Media pătratică simplă se determină cu expresia:

$$m_p = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}} \qquad (5.35)$$

Media pătratică ponderată se determină cu relația:

$$m_p = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 f_i}{\sum_{i=1}^{n} f_i}} \qquad (5.36)$$

Calculul se face în următoarea succesiune: se ridică termenii seriei statistice la pătrat; se împarte suma la numărul termenilor; se extrage rădăcina pătrată din rezultat.

Exemplu de calcul 5.8. Se cere să se determine productivitatea medie pe muncitorii din două ateliere, pe baza înregistrării următoarelor date privind producție de piese pe muncitor din cele două ateliere.

Material cu drept de autor

Media aritmetică $\mathbf{m_a}$ va fi:

Atelierul I: $m_{a1} = \frac{\sum x_i}{n} = \frac{61}{8} = 7,62$ piese.

Atelierul II: $m_{a2} = \frac{\sum x_i f_i}{\sum f_i} = \frac{186}{26} = 7,15$ piese

Pornind de la datele privind producția muncitorilor la cele două ateliere se determină cele două medii pătratice.

Muncitorul	Nr. de piese/ muncitor x_i	Nr. de piese/ muncitor x_i	Nr. de muncitori f_i	$x_i.f_i$
Atelierul I		**Atelierul II**		
A	4	4	3	12
B	6	6	4	24
C	5	5	2	10
D	8	8	5	40
E	7	7	6	42
F	9	9	3	27
G	10	10	2	20
H	11	11	1	11
Total	61		26	186

$$m_{p1} = \sqrt{\frac{\sum x_i^2}{n}} = \sqrt{\frac{4^2 + 6^2 + 5^2 + 8^2 + 7^2 + 9^2 + 10^2 + 11^2}{8}} = \sqrt{\frac{492}{8}} = 7,84$$

$$m_{p2} = \sqrt{\frac{\sum x_i^2 f_i}{\sum f_i}} = \sqrt{\frac{4^2.3 + 6^2.4 + 5^2.2 + 8^2.5 + 7^2.6 + 9^2.3 + 10^2.2 + 11^2}{3+4+2+5+6+3+2+1}} = \sqrt{\frac{1370}{26}} = 7,26$$

Media cronologică **($\mathbf{m_{cr}}$)** este o formă transformată a mediei aritmetice, deci, este o medie generală calculată din medii parțiale. Media cronologică simplă caracterizează tendința de evoluție a seriilor cronologice de momente de intervale egale, iar media cronologică ponderată este utilizată pentru seriile cronologice de momente cu intervale neegale (se prezintă în capitolul 8).

Media cronologică este folosită în statistică pentru determinarea nivelului mediu al seriilor cronologice de momente şi, aşa cum s-a specificat, este de două feluri: simplă şi ponderată.

Media cronologică simplă se întrebuințează în situațiile în care termenii dinamice statistice sunt plasați la distanțe egale unul de altul. Se determină însumând jumătate din valoarea primului şi ultimului termen al seriei dinamice, cu valoarea întreagă a celorlalți termeni şi raportând suma obținută la numărul termenilor minus 1 pe baza expresiei:

Material cu drept de autor

$$m_{cr} = \frac{\frac{x_1}{2} + x_2 + x_3 + \ldots + x_{n-1} + \frac{x_n}{2}}{n-1} \quad (5.37)$$

Exemplu de calcul 5.9. Să se determine valoare medie a fundurilor fixe pentru o societate comercială, anii 1994-2000, dacă la 31 decembrie pentru fiecare an existau valorile:

Anii	1994	1995	1996	1997	1998	1999	2000
u.m.	500	520	540	600	620	650	700

$$m_{cr} = \frac{\frac{500}{2} + 520 + 540 + 600 + 620 + 650 + \frac{700}{2}}{6} = \frac{3530}{6} = 588$$

Media cronologică ponderată se întrebuințează în cazurile în care termenii seriei nu se află la distanță egală unul de altul. În acest caz, semisuma termenilor se ponderează cu distanța dintre momente:

$$m_{cr} = \frac{\frac{x_1 + x_2}{2} \times f_1 + \frac{x_2 + x_3}{2} \times f_2 + \frac{x_3 + x_4}{2} \times f_3 + \ldots + \frac{x_{n-1} + x_n}{2} \times f_{n-1}}{f_1 + f_2 + \ldots + f_{n-1}} \quad (5.38)$$

Exemplu de calcul 5.10. Stocul de mărfuri al unei baze de desfacere, conform inventarului efectuat, a înregistrat următoarele valori în unități monetare (u.m.): 1 ianuarie: 300 u.m.; 1 martie: 400 u.m.; 1 iunie: 450 u.m.; 1 iulie: 500 u.m. Deci $f_1 = 2$ luni; $f_2 = 3$ luni; $f_3 = 1$ lună.

Prin aplicarea formulei de la media cronologică ponderată se obține:

$$m_{cr} = \frac{\frac{300+400}{2} \times 2 + \frac{400+450}{2} \times 3 + \frac{450+500}{2} \times 1}{2+3+1} = \frac{2450}{6} = 408{,}33 \ u.m.$$

Expresiile de calcul ale mediei cronologice reprezintă o adaptare a formulelor mediei aritmetice la elementele seriilor cronologice. Dacă reprezentăm alăturat expresiile de calcul ale celor două medii se obține:

Media	Aritmetică	Cronologică
Simplă	$m_a = \frac{\sum_{i=1}^{n} x_i}{n}$	$m_{cr} = \frac{\frac{x_1}{2} + x_2 + x_3 + \ldots + x_{n-1} + \frac{x_n}{2}}{n-1}$
Ponderată	$m_a = \frac{\sum_{i=1}^{p} x_i n_i}{\sum_{i=1}^{p} n_i}$	$m_{cr} = \frac{\frac{x_1+x_2}{2} \times f_1 + \frac{x_2+x_3}{2} \times f_2 + \ldots + \frac{x_{n-1}+x_n}{2} \times f_{n-1}}{f_1 + f_2 + \ldots + f_{n-1}}$

Se poate demonstra simplu că între diferitele valori medii există următoarele inegalități:

$$M_{-1}(X_i) \le M_0(X_i) \le M(X_i) \le M(X_i^2)$$

Material cu drept de autor

adică:

$$\mathbf{m_h \le m_g \le m_a \le m_p \le m_{cr}} \qquad (5.39)$$

şirul acestor inegalități putând continua la infinit, pentru oricare din numerele puterilor de ordinul **k, k', k'', ...** care satisfac inegalitățile: **k < k' < k''** <.... Va exista astfel totdeauna: $\mathbf{M_k < M_{k'} < M_{k''}}$ ... dacă valorile $\mathbf{X_1, X_2,...,X_n}$ sunt pozitive. S-a notat:

$$M_k = M(X_i^k); M_{k'} = M(X_i^{k'}); M_{k''} = M(X_i^{k''}) etc.$$

Toate aceste medii se folosesc în funcție de specificul şi de proprietățile fenomenului economic, respectând întocmai cerința ca media aleasă să reflecte cât mai fidel tendința fenomenului respectiv.

Media aritmetică evaluată pe baza frecvențelor relative. Frecvențele relative arată importanța relativă a fiecărei grupe, punând în evidență structura colectivității. Folosirea acestui indicator permite să se compare seriile empirice între ele sau cu distribuțiile teoretice.

Se calculează cu relația:

$$n_i^* = \frac{n_i}{\sum_{i=1}^{k} n_i} \Rightarrow n_i^*\% = \frac{n_i}{\sum_{i=1}^{k} n_i} \times 100 \qquad (5.40)$$

unde: $\mathbf{n_i}$ - frecvența absolută a grupei i; n_i^* - frecvența relativă a grupei **i**, exprimată sub formă de coeficient; $n_i^*\%$ - frecvența relativă a grupei **i**, exprimată procentual. Ca urmare:

$$\bar{x} = \frac{\sum x_i n_i^*\%}{\sum n_i} \qquad (5.41)$$

Media caracteristicii alternative. Caracteristica alternativă înregistrează numai două variante posibile de manifestare. De exemplu, candidat admis/respins, produs bun/rebut, plan de producție realizat/nerealizat, persoană căsătorită/necăsătorită etc.

În practică, orice variabilă poate fi exprimată sub formă alternativă dacă se împarte colectivitatea în două grupe după un anumit criteriu. În cazul unei variabile calitative se consideră o modalitate ca fiind forma directă de manifestare a caracteristicii iar toate celelalte modalități ca opusul formei directe de manifestare a caracteristicii. Pentru variabilele numerice cel mai frecvent se utilizează criteriul mediei, structurându-se colectivitatea în unități cu valoarea caracteristicii sub/peste medie.

Alteori, în locul mediei se alege o valoare semnificativă a caracteristicii.

În vederea satisfacerii necesităților de calcul, cele două variante ale caracteristicii alternative pot fi codificate cu:

- "1" - variantă pozitivă, răspunsul afirmativ;

Material cu drept de autor

• "0" - variantă opusă, răspuns negativ.

Pentru calculul mediei aritmetice a caracteristicii alternative utilizează formula mediei ponderate:

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i . n_i}{\sum_{i=1}^{k} n_i} \qquad \text{unde } \mathbf{k = 2}.$$

Conform notațiilor din tabelul următor, formula de mai sus devine:

$$p = \frac{1.M + 0.(N - M)}{N} = \frac{M}{N}, \qquad (5.42)$$

unde: **M** - frecvența (numărul) unităților care au varianta afirmativă; **N** - volumul colectivității statistice.

Variante	$\mathbf{x_i}$	$\mathbf{n_i}$
Afirmativ	**1**	**M**
Negativ	**0**	**N - M**
Total	-	**N**

După cum rezultă din relația (5.42), media caracteristicii alternative este o mărime relativă de structură, care arată ponderea unităților care îndeplinesc o anumită condiție în total colectivitate.

5.5.2. Indicatori de poziție

Din punctul de vedere al analizei statistice, pe lângă informațiile cu caracter sintetic privind valorile individuale, este important şi modul în care sunt repartizate aceste valori.

Mediile se calculează pe baza tuturor valorilor individuale ale seriei, ceea ce le face sensibile la valorile extreme, mai puțin semnificative.

Uneori, valorile extreme ale seriei sunt excesiv de îndepărtate de centrul seriei, ceea ce afectează în mare măsură reprezentativitatea mediei. Alteori, unitățile seriei au tendința de a se concentra la una din extremitățile seriei, rezultând distribuții asimetrice la dreapta sau la stânga.

Pentru aceasta se calculează indicatorii medii de poziție (de structură): **mediana; cuartilele; decilele; modul**.

Mediana. Mediana este acea valoare a unei serii statistice ordonate crescător sau descrescător, care împarte seria în două părți egale. Aşadar, mediana este valoarea din centrul seriei: jumătate din termeni sunt mai mici sau egali cu mediana, jumătate sunt mai mari sau egali. Mediana este o mărime care depinde în primul rând de numărul termenilor.

Pentru *seriile simple*, mediana este termenul din mijloc (dacă seria are număr impar de termeni) sau se determină, în mod convențional, ca

Material cu drept de autor

medie aritmetică simplă a celor două valori din centrul seriei, dacă numărul termenilor este par.

Pentru *seriile de distribuție de frecvențe*, calculul medianei parcurge următoarele etape:

Se identifică grupa în care este inclusă mediana care este acea grupă a cărei frecvență cumulată este prima mai mare decât: $\left(\sum_{i=1}^{k} n_i + 1\right) \Big/ 2$, iar valoarea medianei se calculează cu expresia:

$$Me = x_0 + h \times \frac{\frac{\sum_{i=1}^{k} n_i + 1}{2} - \sum_{i=1}^{m-1} n_i}{n_m} \tag{5.43}$$

în care **h** este mărimea intervalului median; **m** – indexul intervalului median; $\sum_{i=1}^{m-1} n_i$ - suma frecvențelor precedente intervalului median (frecvența cumulată a intervalelor precedente celui median); $\mathbf{n_m}$ – frecvența absolută a intervalului median.

Mediana se exprimă în unitatea de măsură a caracteristicii respective.

Cuartile. Cuartilele sunt acele valori ale caracteristicii care separă seria statistică în patru părți egale.

$$Q_I = x_0 + h \frac{\frac{1}{4}\left(\sum_{i=1}^{k} n_i + 1\right) - \sum_{i=1}^{u-1} n_i}{n_u} \tag{5.44}$$

în care **u** este indexul intervalului care conține $\mathbf{Q_I}$ şi care se determină cu expresia:

$$u(Q_I) = \frac{\sum_{i=1}^{k} n_i + 1}{4} \tag{5.45}$$

$$Q_{II} = Me \tag{5.46}$$

$$Q_{III} = x_0 + h \frac{\frac{3}{4}\left(\sum_{i=1}^{k} n_i + 1\right) - \sum_{i=1}^{v-1} n_i}{n_v} \tag{5.47}$$

în care **v** este indexul intervalului care conține $\mathbf{Q_{III}}$ şi care se determină cu expresia:

Material cu drept de autor

$$\nu(Q_{III}) = \frac{3}{4}\left(\sum_{i=1}^{k} n_i + 1\right) \tag{5.48}$$

Şi cuartilele se pot calcula pentru frecvențe relative.

Decile. Decilele sunt acele valori ale caracteristicii care separă seria statistică în zece părți egale. Se calculează analog cu calculul cuartilelor.

Moda (modul, modulul). Moda (**Mo**) este valoarea care se repetă de cele mai multe ori, motiv pentru care mai este cunoscut în literatura de specialitate şi sub denumirea de **dominanta seriei**. În cazul unei serii de distribuție pe variante, moda este varianta cu frecvența maximă.

În cazul grupării pe intervale moda este intervalul cu frecvența maximă iar valoarea se calculează cu expresia:

$$Mo = x_0 + h \times \frac{\Delta_1}{\Delta_1 + \Delta_2} \tag{5.49}$$

în care x_0 este limita inferioară a intervalului modal; h – mărimea intervalului modal; Δ_1 - diferența dintre frecvența intervalului modal şi a celui precedente; Δ_2 - diferența dintre frecvența intervalului modal şi a celui următor.

5.5.3. Momente

Momente inițiale (simple). Se numeşte moment inițial (simplu) de ordinul **q** al distribuției date, expresia:

$$M_q = \frac{\sum_{i=1}^{k} x_i^q . n_i}{\sum_{i=1}^{k} n_i} \tag{5.50}$$

pentru **q = 1**:

$$M_q = \bar{x} = \frac{\sum_{i=1}^{k} x_i . n_i}{\sum_{i=1}^{k} n_i} \tag{5.51}$$

adică este media aritmetică. Pentru **q = 2**:

Material cu drept de autor

$$M_q = \frac{\sum_{i=1}^{k} x_i^2 . n_i}{\sum_{i=1}^{k} n_i} \tag{5.52}$$

adică este media pătratică.

Momente centrate. Dacă se alege ca origine de calcul al momentului de ordinul **q**, momentul de ordinul întâi, atunci se obține momentul centrat de ordinul **q** care are expresia:

$$m_q = \frac{\sum_{i=1}^{k} \left(x_i - \bar{x}\right)^q . n_i}{\sum_{i=1}^{k} n_i} \tag{5.53}$$

Dacă se dezvoltă expresia $(x_i - \bar{x})^q$, se obține:

$$\left(x - \bar{x}\right)^q = x_i^q - C_q^1 x_i^{q-1} \bar{x} + C_q^2 x_i^{q-2} \bar{x}^2 + \ldots + (-1)^{q-1} C_q^q \bar{x}^q$$

Înmulțind cu n_i şi împărțind la $\sum_{i=1}^{k} n_i$ rezultă:

$$m_q = \frac{\sum_{i=1}^{k} \left(x_i - \bar{x}\right)^q . n_i}{\sum_{i=1}^{k} n_i} = \frac{\sum_{i=1}^{k} x_i^q . n_i}{\sum_{i=1}^{k} n_i} - C_q^1 \frac{\sum_{i=1}^{k} x_i^{q-1} . n_i}{\sum_{i=1}^{k} n_i} \bar{x} +$$

$$+ C_q^2 \frac{\sum_{i=1}^{k} x_i^{q-2} . n_i}{\sum_{i=1}^{k} n_i} \bar{x}^2 + \ldots + \left(-1\right)^q \bar{x}^q$$

sau:

$$m_q = M_q - C_q^1 M_{q-1} M_1 + C_q^2 M_{q-2} M_1^2 + \ldots + (-1)^q M_1^q$$

Dacă se dă lui **q** valorile **1, 2,…,q**, se obțin expresiile momentelor centrate de diferite ordine în funcție de momentele obişnuite:

$$m_1 = M_1 - M_1 = 0$$

Material cu drept de autor

$$m_2 = M_2 - 2.M_1.M_1 + M_1^2 = M_2 - M_1^2$$

$$m_3 = M_3 - 3.M_2.M_1 + 3.M_1.M_1^2 - M_1^3 = M_3 - 3.M_2.M_1 + 2.M_1^3$$

.....

5.6. Indicatori de analiză ai tendinței de împrăştiere a datelor statistice

Indicatori simpli. Servesc pentru a caracteriza gradul de împrăştiere a unităților purtătoare ale caracteristicilor înregistrate. Aceşti indicatori se pot exprima atât în mărimi absolute – folosind aceleaşi unități de măsură ca şi pentru caracteristica studiată – cât şi în mărimi relative, calculate în raport cu valoarea medie.

Amplitudinea absolută a variației (A) se calculează ca diferență între nivelul maxim ($\mathbf{x_{max}}$) şi nivelul minim ($\mathbf{x_{min}}$) al caracteristicii:

Amplitudinea absolută a variației (A) se calculează ca diferență între nivelul maxim ($\mathbf{x_{max}}$) şi nivelul minim ($\mathbf{x_{min}}$) al caracteristicii:

$$A = x_{\max} - x_{\min} \qquad (5.54)$$

Amplitudinea relativă a variației (A%) se exprimă, de regulă, în procente şi se calculează ca raport între amplitudinea absolută a variației (**A**) şi nivelul mediu al caracteristicii ($\mathbf{m_a}$):

$$A\% = \frac{A}{m_a}100 \qquad (5.55)$$

Abaterile individuale absolute (a) se calculează ca diferențe între fiecare valoare înregistrată şi media aritmetică a acestora:

$$a = x_i - m_a \qquad (5.57)$$

Abaterile individuale relative (a%) se calculează raportând abaterile absolute la nivelul mediu al caracteristicii:

$$\frac{a_i}{m_a}100 = \frac{x_i - m_a}{m_a}100 \qquad (5.58)$$

Indicatorii sintetici ai variației. Aceştia sunt: abaterea medie liniară, abaterea medie pătratică, dispersia şi coeficientul de variație.

Abaterea medie liniară ($\bar{d}$) se calculează ca medie aritmetică simplă sau ponderată din abaterile termenilor seriei de la media lor, luate în valoare absolută.

- pentru o serie de frecvențe absolute, abaterea medie liniară se determină cu relația:

Material cu drept de autor

$$\overline{d} = \frac{\sum_{i=1}^{n} |x_i - m_a|}{n} = \frac{1}{n} \sum_{i=1}^{n} |x_i - m_a| \qquad (5.59)$$

respectiv:

$$\overline{d} = \frac{\sum_{i=1}^{k} |x_i - m_a| n_i}{\sum_{i=1}^{k} n_i} = \frac{1}{\sum_{i=1}^{k} n_i} \sum_{i=1}^{n} |x_i - m_a| n_i \qquad (5.60)$$

- pentru o serie de frecvențe relative, exprimate în procente, abaterea medie liniară se determină cu relația:

$$\overline{d} = \frac{\sum_{i=1}^{k} |x_i - m_a| n_i \%}{100} = \frac{1}{100} \sum_{i=1}^{n} |x_i - m_a| n_i \% \qquad (5.61)$$

Dispersia. Dispersia unei caracteristici se notează cu σ^2 şi se calculează cu o medie aritmetică simplă sau ponderală a pătratelor abaterilor termenilor față de media lor. Deci, se mai poate numi şi pătratul mediu al abaterilor termenilor față de media lor. Formulele de calcul sunt:

- pentru o serie simplă:

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - m_a)^2}{n} = \frac{1}{n} \sum_{i=1}^{n} (x_i - m_a)^2 \qquad (5.62)$$

- pentru o serie cu frecvențe absolute:

$$\sigma^2 = \frac{\sum_{i=1}^{k} (x_i - m_a)^2 . n_i}{\sum_{i=1}^{k} n_i} = \frac{1}{\sum_{i=1}^{k} n_i} \sum_{i=1}^{n} (x_i - m_a)^2 . n_i \qquad (5.63)$$

şi aplicând relația de dependență dintre momentele centrate de ordinul 2 şi momentele obişnuite, rezultă:

$$\sigma^2 = M_2 - M_1^2 = \frac{\sum_{i=1}^{n} x_i^2 n}{\sum_{i=1}^{n} n_i} - \left(\frac{\sum_{i=1}^{n} x_i n_i}{\sum_{i=1}^{n} n_i} \right)^2 \qquad (5.64)$$

Dacă substituim $n = \sum_{i=1}^{n} n_i$ se poate obține:

Material cu drept de autor

$$\sigma^2 = \frac{\sum_{i=1}^{n} x_i^2 n_i - \frac{\left(\sum_{i=1}^{n} x_i n_i\right)^2}{n}}{n} \qquad (5.64)$$

Dacă valorile inițiale se iau în raport cu originea de calcul **C** şi se împart la un număr constant **a**, atunci valorile inițiale se pot scrie sub forma:

$$x_i = x_i'.a + C$$

şi expresia de calcul a dispersiei va căpăta forma:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i'a + C)^2 n_i}{\sum_{i=1}^{n} n_i} - \left(\frac{\sum_{i=1}^{n}(x_i'a + C)n_i}{\sum_{i=1}^{n} n_i}\right)^2 = a^2\left[\frac{\sum_{i=1}^{n} x_i'^2 .n_i}{\sum_{i=1}^{n} n_i} - \left(\frac{\sum_{i=1}^{n} x_i' n_i}{n}\right)^2\right] \qquad (5.65)$$

sau:

$$\sigma^2 = a^2\left[\frac{\sum_{i=1}^{n} x_i'^2 .n_i - \frac{\left(\sum_{i=1}^{n} x_i' n_i\right)^2}{n}}{n}\right] \qquad (5.66)$$

- pentru o serie cu frecvențe relative exprimate în procente:

$$\sigma^2 = \frac{\sum_{i=1}^{k}(x_i - m_a)^2 .n_i\%}{100} = \frac{1}{100}\sum_{i=1}^{n}(x_i - m_a)^2 .n_i\% \qquad (5.67)$$

Dispersia reprezintă momentul centrat de ordinul doi şi se poate stabili şi fără să fie necesar să se calculeze în prealabil abaterile individuale ale variantelor de la media lor. Efectuând dezvoltări succesive în formula dispersiei se obține:

- pentru o serie simplă:

$$\sigma^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2 \qquad (5.69)$$

Material cu drept de autor

- pentru o serie de frecvențe:

$$\sigma^2 = \frac{\sum_{i=1}^{k} x_i^2 .n_i}{\sum_{i=1}^{k} n_i} - \left(\frac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i} \right)^2 \qquad (5.70)$$

Abaterea medie pătratică. Este numită şi **abatere standard** sau **abatere tip**, notată cu litera grecească σ, se calculează ca o medie pătratică din abaterile tuturor variantelor seriei de la media lor aritmetică. Este, în fond, radicalul din dispersie:

- pentru o serie simplă:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - m_a)^2}{n}} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - m_a)^2} \qquad (5.71)$$

- pentru o serie de frecvențe absolute:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k} (x_i - m_a)^2 .n_i}{\sum_{i=1}^{k} n_i}} = \sqrt{\frac{1}{\sum_{i=1}^{k} n_i} \sum_{i=1}^{n} (x_i - m_a)^2 .n_i} \qquad (5.72)$$

- pentru o serie cu frecvențe relative exprimate în procente:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k} (x_i - m_a)^2 .n_i\%}{100}} = \sqrt{\frac{1}{100} \sum_{i=1}^{n} (x_i - m_a)^2 .n_i\%} \qquad (5.73)$$

Coeficientul de variație (V). Se calculează în raport cu abaterea medie pătratică şi nivelul mediu al seriei.

$$V = \frac{\sigma}{m_a} 100 \qquad (5.74)$$

Dacă se cunoaşte numai abaterea medie liniară, atunci se poate calcula coeficientul de variație şi pe baza expresiei:

$$V_d = \frac{\bar{d}}{m_a} 100 \qquad (5.75)$$

Material cu drept de autor

Exemplu de calcul 5.11. La o unitate economică cu 80 salariați se face o analiză statistică pentru veniturile anuale ale acestora. Se înregistrează următoarea situație:

Grupe, după venit, u.m.	90-95	95-100	100-105	105-110	110-115	115-120	120-125
Nr. salariați	3	1	42	23	9	1	1

Se cere să se determine indicatorii de variație pentru datele înregistrate mai sus.

Rezolvare. Se întocmeşte următorul tabel de calcul ajutător, care ilustrează sugestiv datele obținute:

Grupe de salariați, după venitul înregistrat, **u.m.**	Numărul salariaților, (n_i)	Centrul intervalului (x_i)	$x_i - m_a$	$\lvert x_i - m_a \rvert n_i$	$(x_i - m_a)^2 n_i$
90-95	3	92,5	-12,6	37,8	476,8
95-100	1	97,5	-7,6	7,6	57,76
100-105	42	102,5	-2,6	109,2	283,92
105-110	23	107,5	2,4	55,2	132,48
110-115	9	112,5	7,4	66,6	492,84
115-120	1	117,5	12,4	12,4	153,76
120-125	1	122,5	17,4	17,4	302,76
Total	80	-	-	306,2	1899,80

În primul rând se determină valoarea medie m_a. În acest scop se întocmeşte tabelul următor:

Grupe de salariați, după venitul înregistrat, **u.m.**	Nr. salariaților, (n_i)	Centrul intervalului, (x_i)	$x_i - a$	$\frac{x_i - a}{k}$	$\frac{x_i - a}{k} n_i$
90-95	3	92,5	-10	-2	-6
95-100	1	97,5	-5	-1	-1
100-105	42	102,5	0	0	0
105-110	23	107,5	+5	1	23
110-115	9	112,5	+10	2	18
115-120	1	117,5	+15	3	3
120-125	1	122,5	+20	4	4
Total	80	-	-	-	41

Se va utiliza astfel o expresie simplificată:

$$m_a = \frac{\sum_{i=1}^{n} \frac{x_i - a}{k}}{\sum_{i=1}^{n} n_i} k + a$$

Material cu drept de autor

Acest procedeu se bazează pe două proprietăți ale mediei aritmetice, şi anume: dacă toți termenii seriei se micşorează cu o constantă **a**, media noilor termeni este mai mică cu acea constantă şi dacă toți termenii se împart cu un coeficient **k**, se va micşora şi ea de acelaşi număr de ori (se adoptă **k = 5** deoarece toți termenii sunt divizibili prin **5** şi **a = 102,5**, coordonata centrală a intervalului cu cea mai mare frecvență absolută):

$$m_a = \frac{\sum_{i=1}^{n} \frac{x_i - a}{k}}{\sum_{i=1}^{n} n_i} k + a = \frac{41}{80} 5 + 102{,}5 = 105{,}06250$$ venitul mediu/salariat

În continuare, se determină indicii de variație:

Abaterea medie liniară:

$$\overline{d} = \frac{\sum_{i=1}^{k} |x_i - m_a| n_i}{\sum_{i=1}^{k} n_i} = \frac{306{,}2}{80} = 3{,}83 \ \ u.m./salariat$$

Dispersia:

$$\sigma^2 = \frac{\sum_{i=1}^{k} (x_i - m_a)^2 . n_i}{\sum_{i=1}^{k} n_i} = \frac{1899{,}80}{80} = 23{,}75$$

Abaterea medie pătratică:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k} (x_i - m_a)^2 . n_i}{\sum_{i=1}^{k} n_i}} = \sqrt{\frac{1899{,}80}{80}} = \sqrt{23{,}75} \approx 4{,}87$$ u.m./salariat

Coeficientul de variație pe baza abaterii medii pătratice:

$$V = \frac{\sigma}{m_a} 100 = \frac{4{,}87}{105{,}1} 100 = 4{,}63\%$$

Coeficientul de variație pe baza abaterii medii liniare:

$$V_d = \frac{\overline{d}}{m_a} 100 = \frac{3{,}83}{165{,}1} . 100 = 3{,}64\%$$

Material cu drept de autor

Observații. Coeficientul de variație pentru medie ia valori între 0 şi 100% conform inegalității: $\mathbf{0 \leq V_m \leq 10}$.

- dacă $\mathbf{V_m = 0}$, înseamnă lipsă de variație, valorile sunt egale între ele şi egale cu media lor, adică $\mathbf{x_1 = x_2 = \ldots = x_n = m_a}$;
- dacă $\mathbf{V_m \to 0}$, variația caracteristicii este mică;
- dacă $\mathbf{V_m \to 100}$, variația caracteristicii este mare.

Intervalul pentru V_m se poate divide astfel:

a. $\mathbf{0 < V_m \leq 35\%}$ = variație mică care se concretizează în:

• media ca indicator al tendinței centrale este semnificativă pentru că abaterile $\mathbf{(x_i - m_a)}$ sunt mici, valorile seriei gravitează în jurul mediei;

• colectivitatea este omogenă, respectiv este formată din unități ce aparțin aceluiaşi tip calitativ;

• gruparea ca metodă de sistematizare primară este bine făcută.

b. $\mathbf{35\% < V_m \leq 50\%} \Rightarrow$ variație relativ mare, ceea ce înseamnă că aspectele menționate la punctul „a" sunt discutabile;

c. $\mathbf{50\% < V_m \leq 100\%} \Rightarrow$ variație foarte mare ceea ce înseamnă că:

• media calculată nu este semnificativă deoarece ascunde abateri mari ale termenilor care se plasează la distanțe mari;

• colectivitatea cercetată este eterogenă deoarece este formată din unități ce aparțin unor tipuri calitative diferite;

• se va reface gruparea unităților.

Coeficientul de variație poate fi folosit şi ca test de semnificație a reprezentativității mediei, considerându-se următoarele praguri de semnificație:

$\mathbf{0 < m_a \leq 17\%}$ media este strict reprezentativă;

$\mathbf{17\% < m_a \leq 35\%}$, media este moderat semnificativă;

$\mathbf{35\% < m_a \leq 50\%}$ media este relativ reprezentativă;

$\mathbf{m_a > 50\%}$ media nu este reprezentativă.

Revenindu-se la exemplul considerat, pe baza rezultatelor obținute se poate afirma că media este o valoare foarte reprezentativă pentru seria din care s-a calculat, deoarece s-a obținut un coeficient foarte mic de variație (sub 5%); iar cele mai multe valori (65 din 80) se concentrează în intervalul 100÷110. Tendința de normalitate a distribuției se poate constata şi pe baza relației dintre abaterea medie pătratică şi abaterea medie liniară. În exemplul considerat 4/5 din abaterea medie pătratică a seriei sunt egale cu $3{,}896\left(\frac{4 \times 4{,}87}{5}\right)$, ceea ce prezintă o diferență minimă față de valoarea reală a abaterii medii liniare (d = 3,86).

Calculul simplificat pentru abaterea medie pătratică. Ca şi în cazul mediei aritmetice, şi pentru abaterea medie pătratică se poate folosi o metodă simplificată. Se întocmeşte tabelul următor:

Material cu drept de autor

Grupe de salariați, după venitul înregistrat, **u.m.**	Nr. salariaților, ($\mathbf{n_i}$)	$\frac{x_i - a}{k}$	$\left(\frac{x_i - a}{k}\right)^2$	$\left(\frac{x_i - a}{k}\right)^2 n_i$
90-95	3	-2	4	12
95-100	1	-1	1	1
100-105	42	0	0	0
105-110	23	1	1	23
110-115	9	2	4	36
115-120	1	3	9	9
120-125	1	4	10	16
Total	80	-	-	97

Dispersia se va determina cu următoarea relație simplificată:

$$\sigma^2 = \frac{\sum_{i=1}^{k}\left(\frac{x_i - a}{k}\right)^2 n_i}{\sum_{i=1}^{k} n_i} k^2 - (m_a - a)^2 \qquad (5.76)$$

La această expresie se ajunge considerând două proprietăți ale dispersiei:

- dispersia calculată față de o constantă **a**

$$\left(\frac{\sum (x_i - a)^2}{n}\right)$$

este mai mică decât dispersia seriei

$$\left(\frac{\sum (x_i - m_a)^2}{n}\right)$$

cu pătratul diferenței dintre medie şi constantă, respectiv $\mathbf{(m_a - a)^2}$.

- Dispersia calculată din simplificarea abaterilor individuale cu un coeficient **k**

$$\left(\frac{\sum (x_i - m_a)}{k}\right)$$

este de $\mathbf{k^2}$ ori mai mică decât dispersia seriei

$$\left(\frac{\sum (x_i - m_a)}{n}\right)^2.$$

Calculată cu expresia simplificată:

Material cu drept de autor

$$\sigma^2 = \frac{97}{80}5^2 - (105{,}1 - 102{,}5)^2 = 23{,}5525.$$

Deci, aproximativ aceeaşi, ca şi în cazul precedent.

5.7. Inegalitatea lui Cebâşev

Matematicianul rus P. L. Cebâşev (1821-1894), fondatorul şcolii matematice din Sankt Petersburg a stabilit o limită pentru probabilitatea de a obține abateri de o mărime dată, printr-o inegalitate care-i poartă numele.

Inegalitatea lui Cebâşev se poate formula în două moduri echivalente:

$$P\left[|X - M(X)| \geq L\right] \leq \frac{\sigma^2}{L^2} \tag{5.77}$$

şi

$$P\left[|X - M(X)| < L\right] \geq 1 - \frac{\sigma^2}{L^2} \tag{5.78}$$

în care **L** este un număr pozitiv oarecare, iar ceilalți indicatori sunt cunoscuți din calculele anterioare.

Inegalitatea (5.77) ne spune că ***probabilitatea ca abaterile în valori absolute să fie mai mari decât un număr pozitiv oarecare L este mai mică decât dispersia împărțită la L^2***.

Acelaşi lucru, însă sub o altă formă, ne spune şi inegalitatea (5.78); în această inegalitate se afirmă că ***probabilitatea ca abaterile în valori absolute să fie mai mici decât un număr pozitiv oarecare L este mai mare decât 1 - σ^2/L^2***. Este vorba deci de o limită inferioară a probabilităților ca abaterile în valori absolute ale unei variabile aleatoare **X** să fie mai mici decât un număr pozitiv oarecare.

Dacă se ia drept unitate de măsură pentru abateri, abaterea medie pătratică, σ, şi se evaluează **L** cu ajutorul acestei unități de măsură, se obține **L = k.**σ, în care **k** este un coeficient oarecare; prin urmare, inegalitatea (5.78) poate fi scrisă şi în felul următor.

$$P\left[|X - M(X)| < k\sigma\right] \geq 1 - \frac{\sigma^2}{k^2\sigma^2} = 1 - \frac{1}{k^2} \tag{5.79}$$

Să dăm lui **k** diferite valori. De exemplu, **k = 2**, avem:

$$P\left[|X - M(X)| < 2\sigma\right] \geq 1 - \frac{\sigma^2}{2^2\sigma^2} = 1 - \frac{1}{2^2} = \frac{3}{4}$$

care înseamnă că – cu o probabilitate cuprinsă între 3/4 şi 1 – majoritatea abaterilor valorilor oricărei variabile aleatoare nu va depăşi **2**σ; adică în medie cel puțin 75% din numărul abaterilor (în valori absolute) vor fi mai mici decât **2**σ.

Material cu drept de autor

5.8. Estimarea parametrilor statistici

5.8.1. Considerații introductive

Plecând de la datele de eşantioane reprezentative, se vor induce rezultate asupra populației mamă (adică populație din care au fost prelevate eşantioanele).

Mai exact, fie θ un parametru necunoscut (în cele ce urmează vom considera că θ este un scalar dar raționamentele pot fi extinse asupra estimării unui vector de parametri, de exemplu vectorul medie sau vectorul varianță care definesc o lege normală) care intervin în legea de probabilitate a unei variabile aleatoare **X**. Legea de probabilitate a acestei variabile aleatoare trebuie să fie cunoscută analitic (se alege printre modelele existente legea cea mai potrivită fenomenului observat). Numai valoarea numerică a parametrului θ care intervine în această lege de probabilitate este necunoscut.

Fie $\mathbf{x_1}, \mathbf{x_2},\ldots,\mathbf{x_n}$, ca valori luate pentru variabila aleatoare **X** într-un eşantion de volum **n** prelevat din populația mamă de volum **N**.

Se numeşte estimator al lui θ, şi se notează cu $\mathbf{T_n}$, funcția care la valorile $\mathbf{x_i}$ a eşantionului face să corespundă valoarea parametrului θ. Se notează valoarea numerică a acestei estimări prin:

$$\hat{\theta} = T_n(x_1, x_2, \ldots, x_n) \tag{5.80}$$

Prin definiție, $\mathbf{T_n}$ este o funcție de realizările unei variabile aleatoare sau, cu alte cuvinte, $\mathbf{T_n}$ este o variabilă aleatoare căruia i se poate încerca să se determine caracteristicile (lege, densitate de probabilitate, funcție de repartiție, Momente etc.).

Exemplu de calcul 5.12: Se observă un fenomen de producție pentru piese manufacturate. Fiecărei piese îi este asociată o măsură (un indicator de calitate). Cum nu se poate verifica fiecare măsură se procedează la o selecție care ne furnizează un eşantion. Presupunem că cunoaşterea naturii acestui indicator ne permite să facem ipoteza că el ascultă de o lege de probabilitate normală. Problema este acum, în ceea ce priveşte eşantionul $\{\mathbf{x_i}\}$, să se propună o valoare pentru media acestei legi normale. Trebuie să procedăm la o estimare a parametrului adevărat μ care se traduce prin $\hat{\mu}$. Există o infinitate de feluri posibile printre care se pot cita:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \; ; \; \hat{\mu} = mediana\{x_i\} ; \; \hat{\mu} = \mathrm{mod}\, ul\{x_i\} ; \; \hat{\mu} = x_7$$

Care este cel mai bun estimator al mediei ? Există el ?

Cu acest exemplu simplu este rezumată problema fundamentală a estimării: care este definiția matematică al celui mai bun ?

Răspunsul este simplu, el nu există. Atunci cum să comparăm estimatorii. Pentru aceasta ne servim de mai multe criterii, cel mai adesea legate de bunul simț:

Material cu drept de autor

Deplasarea: se doreşte ca estimarea să nu fie sistematic decalată în raport cu valoarea adevărată.

Precizia: dacă se repetă estimarea asupra unui alt eşantion, se doreşte obținerea unei estimări coherente, deci să fie mică variația de la un eşantion la altul. Se va vorbi de asemenea de eficacitate.

Convergența: dacă se poate estima valoarea parametrului pe o întreagă populație mamă, valoarea estimației obținute trebuie să fie valoarea adevărată a parametrului.

Complexitatea: orice estimație necesită un calcul, deci un timp. Vom urmări deci să evaluăm complexitatea calculului în funcție de volumul datelor (adică de **n**).

Robustețea: În orice caz concret, există surse de perturbație. Se doreşte ca estimarea să nu fie sensibilă de prezența valorilor accidentale.

Aceste diferite criterii nu sunt neapărat compatibile între ele şi se regăsesc dilemele clasice: precizie sau robustețe, convergență sau complexitate.

5.8.2. Estimarea prin interval de încredere

Există mai multe metode privind teoria şi practica estimării parametrilor statistici. În cele ce urmează se va prezenta numai metoda prin intervale de încredere.

Această metodă de estimație este preferată în practică pentru că ea introduce noțiunea de incertitudine. Se încearcă determinarea intervalului [a,b] centrat pe valoarea numerică estimată a parametrului necunoscut θ conținând valoarea adevărată cu o probabilitate α fixată apriori. Această probabilitate permite adaptarea la exigențele aplicației.

$$P(a < \theta < b) = \alpha \tag{5.81}$$

Intervalul [a,b] este numit interval de încredere şi α coeficientul de încredere.

O estimare prin interval de încredere va fi cu atât mai bună cu cât intervalul va fi mai mic pentru un coeficient de încredere mare.

Datele inițiale, în afară de eşantion, va fi cunoaşterea legii de probabilitate a parametrului de estimat. Cum nu există o rezolvare generală a acestei probleme, vom aborda succesiv cazurile cele mai frecvente (estimarea unei proporții, a unei medii, a unei variațe a legii normale).

5.8.3. Estimarea unei proporții

Fie o populație a cărei indivizi posedă un caracter **A** cu o probabilitate **p** (legea **0/1**). Se caută să se determine această probabilitate necunoscută prelevând un eşantion de volum **n** din această populație. Se constată că un număr **x** dintre cei **n** indivizi posedă caracterul **A**. Ce putem deduce din aceasta, adică proporția $\mathbf{f_n = x/n}$ aproximează valoarea adevărată **p**, dar cu ce încredere.

Material cu drept de autor

Fie $\mathbf{F_n = x/n}$; $\mathbf{F_n}$ este o variabilă aleatoare construită prin suma a **n** variabile aleatoare **0/1** şi de acelaşi parametru **p**. Este deci, după teorema central limită, o v.a. a cărei lege de probabilitate tinde către o lege normală de medie **p** şi de abatere tip $\sqrt{[p.(1-p)]/n}$ Această aproximație este valabilă numai dacă volumul eşantionului este suficient de mare (adică **n>30** în practică).

Să construim intervalul de încredere în jurul lui **p** sub forma:

$$P(|f_n - p| < t) = 1 - \alpha \tag{5.82}$$

unde α este riscul (apriori, se construieşte un interval simetric); $\mathbf{f_n}$ este o realizare a unei variabile aleatoare $N(p, \sqrt{[p.(1-p)]/n})$. Deci, se poate, prin normalizare şi centrare, să se obțină o nouă variabilă aleatoare **u**.

$$u = \frac{f_n - p}{\sqrt{[p.(1-p)]/n)}} : N(0,1) \tag{5.83}$$

De aici se deduce intervalul de încredere sub forma:

$$P(a < \theta < b) = P[f_n - u.\sqrt{[p.(1-p)]/n)} < p < f_n + u.\sqrt{[p.(1-p)]/n)}] = 1 - \alpha \tag{5.84}$$

Valoarea $t = u.\sqrt{[p.(1-p)]/n)}$ este deci un rezultat de calcul. Valoarea lui **u** va fi citită într-un tabel a legii normale **N(0, 1)**. Există de altfel diferite moduri de aproximare a valorii lui **p**:

- Fie prin proporția $\mathbf{f_n}$:

$$P(a < \theta < b) = P[f_n - u.\sqrt{[p.(1-p)]/n)} < p < f_n + u.\sqrt{[p.(1-p)]/n)}] = 1 - \alpha$$

- Fie prin majorare. Într-adevăr, oricare ar fi valoarea lui **p**, produsul **p.(1-p)** este majorat cu **1/4**.

$$P(a < \theta < b) = P\left[f_n - \frac{u}{2.\sqrt{n}} < p < f_n + \frac{u}{2.\sqrt{n}}\right] \geq 1 - \alpha$$

Exemplu de calcul 5.13. Fie un eşantion de volum **n = 100** şi o proporție estimată $\mathbf{f_n = 0,6}$. Care este încrederea în această valoare sau ce interval dă o încredere de **0,9** (risc de **10%**) ?

$$t : P\left(0,6 - u.\sqrt{\frac{0,6 \times 0,4}{100}} < p < 0,6 + \sqrt{\frac{0,6 \times 0,4}{100}}\right) = P\left(-t < \frac{p - 0,6}{0,049} < t\right) = 0,9$$

Prin citirea în tabelul legii normale, se obține $P(X < u) = 0,95 \Rightarrow u = 1,645$. Intervalul de **90%** încredere, în jurul proporției estimate, este deci [**0,5194; 0,6808**].

Material cu drept de autor

5.8.4. Estimarea unei medii

Sunt două cazuri de considerat:

- variabila aleatoare măsurată este normală şi număr de realizări este oarecare.

- variabila aleatoare măsurată nu este normală şi numărul de realizări este mai mare decât 30 (în acest caz distribuția mediei tinde către o lege normală, după teorema central limită).

Fie deci o variabilă aleatoare **X** care urmează o lege normală de medie μ necunoscută şi de abatere tip σ. Se dispune de un eşantion de **n** realizări $\mathbf{x}_i$ a acestei variabile aleatoare. Ca şi mai înainte, intervalul de încredere pentru medie este:

$$P(a<\mu<b)=P\left[m-z_\alpha\frac{\sigma}{\sqrt{n}}<\mu<m+z_\alpha\frac{\sigma}{\sqrt{n}}\right]=\alpha \qquad (5.85)$$

Unde **m** este media aritmetică calculată plecând de la eşantion. Pentru a merge mai departe, trebuie să considerăm două cazuri:

1. Varianța σ^2 este cunoscută.

Valoarea σ joacă rolul unei constante în formula intervalului de încredere şi noua variabilă aleatoare $Y=[(m-\mu).\sqrt{n}]/\sigma$ urmează tot o lege normală. Valoarea lui $\mathbf{z}_\alpha$ este deci citită într-o tabelă a legii normale.

2. Varianta σ^2 este necunoscută.

În acest caz, σ joacă rolul unei variabile aleatoare. Fie $\mathbf{s}^2$ estimarea lui σ^2 care se obține prin relația:

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-m)^2 \qquad (5.86)$$

Cum **X** urmează o lege normală, se ştie că cantitatea $(n.s^2)/\sigma^2$ urmează o lege χ^2 cu **n-1** grade de libertate. Noua variabilă aleatoare $Y=[(m-\mu).\sqrt{n}]/s$ urmează deci o lege **t** a lui Student cu **n-1** grade de libertate. Intervalul de încredere este atunci:

$$P(a<\mu<b)=P\left[m-t\frac{s}{\sqrt{n}}<\mu<m+t\frac{s}{\sqrt{n}}\right]=\alpha \qquad (5.87)$$

Unde **t** este citit în tabelul Student pentru **n-1** grade de libertate.

Aposteriori, poate interesa volumul minim al eşantionului astfel ca intervalul de încredere, pentru un coeficient de încredere α dat, să fie aşa fel încât bornele sale interioare şi superioare să nu se depărteze mai mult de k% din valoarea medie. Se impune deci $\left(t.\sigma\right)/\sqrt{n}\leq k.\mu$ ceea ce conduce la:

Material cu drept de autor

$$n \geq \frac{t^2}{k^2}\left(\frac{\sigma}{\mu}\right)^2 \qquad (5.88)$$

Se aproximează μ prin **m** şi σ prin **s** dacă abaterea tip este necunoscută.

5.8.5. Estimarea unei varianțe

Nu vom aborda decât cazul estimării varianței σ^2 a unei variabile aleatoare **X** normală de medie μ plecând de la un eşantion de **n** valori.

Dacă μ este cunoscută (foarte rar), atunci intervalul de încredere cu α**%** (risc) este definit prin:

$$\left[\frac{n.\nu}{\chi^2_{1-\alpha/2}(n)}; \frac{n.\nu}{\chi^2_{\alpha/2}(n)}\right] \qquad (5.89)$$

cu $\nu = \frac{1}{n}\Sigma(x_i - \mu)^2$ şi unde $\chi^2_{1-\alpha/2}(n)$ şi $\chi^2_{\alpha/2}(n)$ sunt cuantile de ordin **1-**α**/2** şi α**/2** din legea χ^2 cu **n** grade de libertate.

Dacă μ este necunoscută. Cantitatea $(n.s^2)/\sigma^2$ definită anterior urmează o lege χ^2 cu **n-1** grade de libertate. Intervalul de încredere cu α% (risc) este definit prin:

$$\left[\frac{n.s^2}{\chi^2_{1-\alpha/2}(n-1)}; \frac{n.s^2}{\chi^2_{\alpha/2}(n-1)}\right] \qquad (5.90)$$

unde $\chi^2_{1-\alpha/2}(n-1)$ şi $\chi^2_{\alpha/2}(n-1)$ sunt cuantile de ordinul **1-**α**/2** şi α**/2** din legea χ^2 cu **n-1** grade de libertate.

Se obține rezultatul următor:

$$P(\chi^2 < k) = P[(n.s^2)/\sigma^2 < k] = \alpha \qquad (5.91)$$

(atenție, α reprezintă aici încrederea) **k** este citit pe o tabelă de χ^2 pentru n-1 grade de libertate, de unde rezultă:

$$P\left(\sigma^2 > \frac{n.s^2}{k}\right) = \alpha \text{ cu } s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2 \qquad (5.92)$$

Material cu drept de autor

5.9. Testarea ipotezelor statistice

Un test este un mecanism care permite tranşarea între două ipoteze în vederea rezultatelor unui eşantion. În cazurile care ne interesează, aceste ipoteze vor conduce la estimări (valoarea unui moment, egalitatea variantelor, natura unei legi de probabilitate...). Fie $\mathbf{H_0}$ şi $\mathbf{H_1}$ aceste două ipoteze, dintre care una şi numai una este adevărată. Decizia va ajunge la a alege $\mathbf{H_0}$ sau $\mathbf{H_1}$. Există deci patru cazuri posibile a căror probabilităţi sunt rezumate în tabelul următor:

	H_0 adevărată	H_1 adevărată
Decide H_0	$1 - \alpha$	β
Decide H_1	α	$1 - \beta$

unde α şi β sunt erori de speţa I-a respectiv speţa II-a; α este probabilitatea de a decide $\mathbf{H_1}$ atunci când $\mathbf{H_0}$ este adevărată; β este probabilitatea de a decide $\mathbf{H_0}$ atunci când $\mathbf{H_1}$ este adevărată.

Aceste două erori sunt antagoniste, cu cât α va fi mai mare (respectiv mic), cu atât β va fi mai mic (respectiv mare).

Faptul de a impune un α mic conduce la o regulă de decizie mai strictă care ajunge cel mai adesea la neabandonarea ipotezei $\mathbf{H_0}$ decât în cazuri rarisime şi deci la conservarea acestei ipoteze uneori pe nedrept. Compromisul între valorile lui α şi β este de dorit dar foarte greu de realizat.

Se numeşte puterea unui test cantitatea $\mathbf{1}$-β.

În practica testelor statistice, este obişnuit a se fixa α dată (valorile cele mai curente sunt 0,05, 0,01, 0,1) de preferinţă în funcţie de riscul de speţa I-a. Într-adevăr, $\mathbf{H_0}$ joacă cel mai adesea un rol predominant în raport cu ipoteza $\mathbf{H_1}$. Aceasta este consecinţă a faptului că $\mathbf{H_0}$ joacă rolul de ipoteză de referinţă, în timp ce $\mathbf{H_1}$ este adesea limitată la ipoteza contrarie. De exemplu, putem avea $\mathbf{H_0}$: $\mathbf{m = m_0}$ ceea ce este relativ uşor de testat şi în acest caz $\mathbf{H_1}$ este pur şi simplu $\mathbf{m \neq m_0}$.

Această practică este legată de faptul că evaluarea unui test trece prin evaluarea funcţiilor complexe care au fost tabelate pentru numeroase valori ale lui α dar nu sunt cunoscute chiar pentru orice α. Suntem deci conduşi la alegerea apriori a lui α. Totuşi, apariţia din ce în ce mai frecventă a proceselor numerice de aproximare rapidă şi precisă permite o altă aproximare constând în căutarea celei mai mici valori ale lui α pentru care ipoteza valorii $\mathbf{H_0}$ rămâne adevărată.

5.9.1. Teste unilaterale şi bilaterale

Înainte de a aplica orice test statistic, este vorba de definirea problemei puse. Într-adevăr, după ipotezele formulate, se aplică fie un test bilateral, fie un test unilateral.

Material cu drept de autor

Un test bilateral se aplică când se caută o diferență între cele două estimări, sau între o estimare şi o valoare dată, fără a se preocupa de semn sau se sensul diferenței. În acest caz, zona de eliminare (conform paragrafului următor) a ipotezei principale se face de o parte şi de alta a distribuției de referință.

Un test unilateral se aplică când se caută să ştim dacă o estimare este superioară (mică, sau inferioară) unei alteia sau unei valori date. Zona de excludere a ipotezei principale este situată de o singură parte a distribuției de probabilitate de referință.

Anumite teste ca analiza variației sau testul χ^2 (hi pătrat) sunt practic unilaterale.

5.9.2. Regiune de acceptare şi regiune critică

Care este demersul general ? α fiind fixat, trebuie să alegem o variabilă de decizie, variabilă care trebuie să aducă informație asupra problemei puse, conform alegerii între cele două ipoteze. Legea acestei variabile trebuie să fie perfect cunoscută cel puțin într-o ipoteză (cel mai adesea H_0) pentru a nu introduce necunoscute noi în problemă. Se numeşte atunci regiune critică şi se notează cu υ, mulțimea valorilor variabilei de decizie care conduc la îndepărtarea lui H_0 în profitul lui H_1. Se poate lega **W** de alfa prin relația: $P(W|H_0) = \alpha$.

Se numeşte regiune de acceptare şi se notează $\overline{W}$ regiunea complementară regiunii critice. Avem de asemenea relații cu erori de prima şi a doua speță: : $P(\overline{W}|H_0) = 1 - \alpha$. şi $P(W|H_1) = 1 - \beta$. Zona de acceptare corespunde intervalului în care diferențele observate între realizări şi teorie sunt atribuibile fluctuațiilor de eşantionare. Regiunea critică sau zona de excludere corespunde deci intervalelor în care diferențele sunt prea mari pentru a fi fructul hazardului eşantionării.

Construcția unui test este determinarea apriori a regiunii critice fără a cunoaşte rezultatul experienței. Se poate rezuma acest demers în felul următor.

- alegerea lui H_0 şi H_1,
- determinarea variabilei de decizie;
- aliura regiunii critice în funcție de H_1;
- calculul regiunii critice în funcție de α;
- calculul eventual al puterii testului $1 - \beta$;
- calculul experimental al variabilei de decizie;
- concluzia testului: excluderea sau acceptarea lui H_0.

5.9.3. Alegerea unui test

Multe teste de concepție foarte diferită sunt adesea disponibile pentru a supune unei probe de adevăr o ipoteză principală. Într-un anume

Material cu drept de autor

caz, testul care furnizează eroarea β cea mai mică pentru aceeaşi valoare a lui α este prin definiție cel mai puternic (acela care are cea mai mare valoare a puterii testului **1** - β). Într-adevăr, el poate detecta cele mai mici diferențe între populații fără ca pentru aceasta să mărească eroarea de primă speță. Majoritatea testelor statistice se bazează pe respectul unui anumit număr de condiții. După gradul de respectare a acestor condiții de aplicare, validitatea rezultatelor sunt mai mult sau mai puțin afectate şi ea este cu atât mai mare cu cât testul este mai puțin robust. Astfel, robustețea unui test echivalează cu toleranța sa față de respectarea condițiilor de aplicație.

Dacă statisticianul dispune de mai multe teste pentru a verifica o ipoteză, el va alege desigur pe cel mai puternic şi pe cel mai robust.

Testele puțin puternice măresc probabilitatea de a comite o eroare de speța doua. Ori, această eroare poate să se dovedească deosebit de gravă. Într-adevăr, în medicină de exemplu, o analiză care ar clasa ca bolnav un individ sănătos poate avea consecințe tot aşa de grave ca o analiză care ar clasa ca sănătoşi indivizii bolnavi (eroare de prima speță). În astfel de cazuri, există un interes în a trasa curba puterii testului. De asemenea numită curbă caracteristică de eficacitate care indică probabilitatea de a lua o decizie bună dacă $\mathbf{H_1}$ este adevărată. Puterea este măsurată prin valoarea **1** - β pentru un α dat.

5.9.4. Influența eşantionării

Pentru a compara mediile, variațele sau alți parametri estimați a două eşantioane, trebuie să luăm în considerare tehnica ce conduce la constituirea celor două eşantioane. Dacă selecția elementelor este aleatoare şi dacă alegerea elementelor din primul eşantion n-are nici o influență asupra alegerii elementelor din al doilea, cele două eşantioane se numesc independente.

Dacă se prelevă aleator perechi de elemente, şi nu elementele însele, se constituie două eşantioane perechi. În acest caz, primul element din fiecare pereche aparține primului eşantion şi cel de al doilea celui de al doilea eşantion. Câteodată, perechea de elemente poate să se raporteze la acelaşi individ la care se măsoară aceeaşi variabilă în două ocazii diferite, de exemplu prin două mijloace diferite.

În ceea ce urmează vom aborda câteva teste clasice. Această listă nu se vrea exhaustivă, în lucrările mai specializate există o aproximație mai sistematică a testelor statistice.

5.9.5. Testul mediei unei legi normale de abatere tip cunoscută

Fie **X** o variabilă aleatoare normală de medie **m** şi de abatere tip σ cunoscută. În vederea unui eşantion de **n** realizări independente $\mathbf{x_i}$ se doreşte să se ştie dacă media **m** este egală cu $\mathbf{m_0}$ sau cu $\mathbf{m_1}$ ceea ce se rezumă prin:

Material cu drept de autor

$$\begin{cases} H_0 : m = m_0 \\ H_1 : m = m_1 \end{cases} \tag{5.93}$$

Funcțiile de veridicitate, sau densitate, a eşantionului sunt:

$$L(x, m_0) = \left(\frac{1}{\sqrt{2.\pi.\sigma}} \right)^n e^{\left[-\frac{1}{2.\sigma^2} \sum_{i=1}^{n} (x_i - m_0)^2 \right]} \tag{5.94}$$

$$L(x, m_1) = \left(\frac{1}{\sqrt{2.\pi.\sigma}} \right)^n e^{\left[-\frac{1}{2.\sigma^2} \sum_{i=1}^{n} (x_i - m_1)^2 \right]} \tag{5.95}$$

Regiunea critică este definită prin raportul acestor două funcții. Trecând printr-un operator logaritm se obține uşor:

$$\sum_{i=1}^{n} (x_i - m_1)^2 - \sum_{i=1}^{n} (x_i - m_0)^2 \le 2.\sigma^2.Ln(k) \tag{5.96}$$

Considerând $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ se obține:

$$\left(\overline{x} - \frac{m_0 + m_1}{2} \right).(m_0 - m_1) \le \frac{\sigma^2.Ln(k)}{n} \tag{5.97}$$

Dacă $m_0 < m_1$, se ajunge la:

$$\overline{x} \ge \frac{m_0 + m_1}{2} - Ln(k) \frac{\sigma^2}{n.(m_1 - m_0)} = \lambda \tag{5.98}$$

Regiunea critică este deci definită prin inegalitatea $\overline{x} \ge \lambda$ pe care acum trebuie să o determinăm. Pentru aceasta, introducem eroarea α. Această eroare este definită prin: $\alpha = P(H_1 | H_0)$. Decidem $\mathbf{H_1}$ dacă $\overline{x} \ge \lambda$, deci $\alpha = P(\overline{X} \ge \lambda | H_0)$. unde $\overline{X}$ este variabila aleatoare a cărei $\overline{x}$ este o realizare. **X** fiind o variabilă aleatoare normală, distribuția lui $\overline{X}$ este de asemenea normală de medie **m** şi de abatere tip $\sigma / \sqrt{n}$.

Avem atunci (condiția $\mathbf{H_0}$ fiind adevărată):

$$\alpha = P(\overline{X} \ge \lambda) \text{ cu } \overline{X} : N\left(m, \frac{\sigma}{\sqrt{n}} \right) \tag{5.99}$$

$$\alpha = P\left(\frac{\overline{X} - m_0}{\sigma / \sqrt{n}} \ge \frac{(\lambda - m_0).\sqrt{n}}{\sigma} \right) \tag{5.100}$$

Material cu drept de autor

Cantitatea $Y = \frac{\overline{X} - m_0}{\sigma/\sqrt{n}}$ urmează o lege normală centrată redusă deci:

$$\alpha = P\left(Y \geq \frac{(\lambda - m_0).\sqrt{n}}{\sigma} \right) \quad \text{cu} \quad Y : N(0,1) \qquad (5.101)$$

Dacă valoarea lui α este fixată, prin citire într-o tabelă a legii normale se poate găsi valoarea $\frac{(\lambda - m_0).\sqrt{n}}{\sigma}$ şi deci celei a lui λ.

Regula de decizie a testului este deci:

Dacă $\overline{x} > \lambda$ se decide H_1 dacă nu, se acceptă H_0.

Printr-un raționament echivalent, se poate evalua eroarea de speța II-a şi deci puterea testului:

$$\beta = P(\overline{X} < \lambda | H_1) = P\left(Y < \frac{(\lambda - m_1).\sqrt{n}}{\sigma} \right) \qquad (5.102)$$

Cu $Y = \frac{\overline{X} - m_1}{\sigma/\sqrt{n}}$ variabilă aleatoare normală centrată redusă.

5.9.6. Testul mediei unei legi normale de abatere tip necunoscută

Raționamentul precedent se aplică până la determinarea lui λ:

$$\alpha = P\left(\frac{\overline{X} - m_0}{s/\sqrt{n}} \geq \frac{(\lambda - m_0).\sqrt{n}}{s} \right) \qquad (5.103)$$

Unde **s** este estimarea abaterii tip necunoascută σ.

Cantitatea $Y = \frac{\overline{X} - m_0}{s/\sqrt{n}}$ nu mai urmează o lege normală centrată redusă căci numitorul nu mai este o constantă ci o realizare a estimatorului de varianță a variabilei **X**. Abaterea standard **s** este obținută prin:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2 \qquad (5.104)$$

Prin construcție $\mathbf{s^2}$ urmează o lege χ^2 şi **Y** va fi deci o variabilă aleatoare care urmează o lege t a lui Student cu **n-1** grade de libertate, ceea ce ne dă:

Material cu drept de autor

$$\alpha = P\left(Y \geq \frac{(\lambda - m_0).\sqrt{n}}{\sigma} \right) \tag{5.105}$$

cu **Y** urmând legea **t** a lui Student cu **n-1** grade de libertate.

Şi aici este posibil ca folosind o tabelă a legii **t** a lui Student să se găsească valoarea pragului de probabilitate şi deci cea a lui α. Regula de decizie este tot aceeaşi.

La fel, printr-un raționament analog, se ajunge la eroarea de speța II-a şi la puterea testului.

$$\beta = P(\overline{X} > \lambda | H_1) = P\left(Y < \frac{(\lambda - m_1).\sqrt{n}}{s} \right) \tag{5.106}$$

cu $Y = \dfrac{\overline{X} - m_1}{s/\sqrt{n}}$ variabila aleatoare Student cu **n-1** grade de libertate.

5.9.7. Test a unei varianțe de lege normală, media fiind cunoscută

Fie **X** o variabilă aleatoare de medie m cunoscută. Se presupune că abaterea tip necunoscută nu poate lua decât două valori: σ_0 şi σ_1. În vederea unui eşantion de **n** realizări independente $\mathbf{x}_i$, se doreşte să se ştie dacă varianța σ^2 este egală cu σ_0^2 sau σ_1^2, ceea ce se rezumă prin:

$$\begin{cases} H_0 : \sigma = \sigma_0 \\ H_1 : \sigma = \sigma_1 \end{cases} \tag{107}$$

Estimatorul varianței va fi:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2 \tag{108}$$

(se utilizează **1/n** şi nu **1/(n-1)** întrucât media este cunoscută.)

Funcțiile de veridicitate sau densitate a eşantionului sunt:

$$L(x, \sigma_0) = \left(\frac{1}{\sqrt{2.\pi.\sigma_0}} \right)^n e^{\left[-\frac{1}{2.\sigma_0^2}\sum_{i=1}^{n}(x_i - m)^2 \right]} \tag{5.109}$$

$$L(x, \sigma_1) = \left(\frac{1}{\sqrt{2.\pi.\sigma_1}} \right)^n e^{\left[-\frac{1}{2.\sigma_1^2}\sum_{i=1}^{n}(x_i - m)^2 \right]} \tag{5.110}$$

Material cu drept de autor

Regiunea critică este definită prin raportul celor două funcții. Trecând printr-un operator logaritmic, se obține uşor:

$$n.Ln\left(\frac{\sigma_1}{\sigma_0}\right)+\frac{1}{2}\sum_{i=1}^{n}(x_i-m)^2\left(\frac{1}{\sigma_1^2}-\frac{1}{\sigma_0^2}\right)\le Ln(k_\alpha) \qquad (5.111)$$

În cazul în care $\sigma_1 > \sigma_0$, se obține:

$$\frac{n.s^2}{\sigma_0^2}\ge\frac{2.\sigma_1^2}{\sigma_0^2-\sigma_1^2}\left(Ln(k_\alpha)-n.Ln\left(\frac{\sigma_1}{\sigma_0}\right)\right) \qquad (5.112)$$

Valoarea $\mathbf{k}_\alpha$ este determinată plecând de la eroarea de speța I-a. Cantitatea $n.s^2/\sigma_0^2$ urmează o lege χ^2 cu **n** grade de libertate. Valoarea prag va fi deci citită într-o tabelă χ_n^2.

5.9.8. Testul unei varianțe de lege normală, media fiind necunoscută

Este cazul mai frecvent decât cel precedent. Tot prin raționament similar, se ajunge le rezultatele următoare:

Variabila de decizie este: $s^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ care este obținută astfel încât $n.s^2/\sigma^2$ urmează o lege χ^2 cu **n-1** grade de libertate. Regiunea critică este definită de $\mathbf{s^2} > \mathbf{k}$ şi **k** este determinat prin:

$$P(s^2>k)=P\left(\chi_{n-1}^2>\frac{n.k}{\sigma_0^2}\right)=\alpha \qquad (5.113)$$

Regula de decizie a testului este deci:
Dacă $\mathbf{s^2} > \mathbf{k}$ se decide $\mathbf{H_1}$, dacă nu, se acceptă $\mathbf{H_0}$.

5.9.9. Testul unei proporții

Fie o populație foarte mare unde proporția de indivizi care au caracterul **A** = **p**. Se crede că această proporție nu poate să aibă decât două valori: $\mathbf{p_0}$ sau $\mathbf{p_1}$. În vederea unui eşantion de volum **n** se doreşte luarea unei decizii cu privire la valoarea acestei proporții, cu semnificația α.

Plecând de la eşantion estimatorul teoretic al proporției va fi frecvența empirică $\mathbf{f_n=x/n}$ unde **x** este numărul de indivizi care au caracterul **A** în eşantion.

Material cu drept de autor

Ipotezele sunt deci:

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p = p_1 \end{cases}$$

Regula deciziei este dată de:

- dacă $\mathbf{f_n \geq \pi}$ atunci $\mathbf{H_1}$;
- dacă $\mathbf{f_n < \pi}$ atunci $\mathbf{H_0}$.

Unde π desemnează regiunea critică.

$\mathbf{f_n}$ este o realizare a unei variabile aleatoare $\mathbf{F_n}$ a cărei lege de probabilitate poate să fie determinată prin teorema central limită. Dacă volumul eşantionului este suficient de mare (în practică $\mathbf{n > 30}$), se admite că legea lui $\mathbf{F_n}$ tinde către o lege normală de medie **p** şi abatere tip $\sqrt{[p.(1-p)]/n}$ ceea ce ne conduce la: $\alpha = \mathbf{P(F_n \geq \pi | H_0)}$ adevărată, cu $\mathbf{F_n}$: $\mathbf{N[p, \sqrt{[p.(1-p)]/n}\,]}$

Sub ipoteza $\mathbf{H_0}$ se obține:

$$\alpha = P\left[\frac{(F_n - p_0).\sqrt{n}}{\sqrt{p_0.(1-p_0)}} \geq \frac{(\pi - p_0).\sqrt{n}}{\sqrt{p_0.(1-p_0)}}\right] = P\left[Y \geq \frac{(\pi - p_0).\sqrt{n}}{\sqrt{p_0.(1-p_0)}}\right] \quad (5.114)$$

unde $Y = [(F_n - p_0).\sqrt{n}]/\sqrt{p_0.(1-p_0)}$ este o variabilă aleatoare normală centrată redusă.

Valoarea de prag critic este citită într-o tabelă a legii normale.

Eroarea de speța II-a şi puterea testului sunt date de:

$$\beta = P\left[Y \leq \frac{(\pi - p_1).\sqrt{n}}{\sqrt{p_1.(1-p_1)}}\right] \quad (5.115)$$

unde $Y = [(F_n - p_1).\sqrt{n}]/\sqrt{p_1.(1-p_1)}$ este o variabilă aleatoare normală centrată redusă.

5.9.10. Test între ipoteze compuse

A. Testul unei medii de lege normală, abaterea tip fiind cunoscută

Test unilateral. Fie **X** o variabilă aleatoare normală de medie **m** şi varianță σ^2 cunoscută. În vederea testării unui eşantion de **n** realizări independente $\mathbf{x_i}$, vrem să alegem între două ipoteze:

Material cu drept de autor

$$\begin{cases} H_0 : m = m_0 \\ H_1 : m < m_0 \end{cases} \quad (5.116)$$

Ca totdeauna, eroarea de speța I-a α este fixată. Media **m** va fi estimată prin media aritmetică $\overline{x}$. Construcția testului este similară cu ceea ce am văzut pentru cazul testului simplu a unei medii. Se ajunge la:

$$\alpha = \left[\frac{(\overline{X} - m_0).\sqrt{n}}{\sigma} \le \frac{(\lambda - m_0).\sqrt{n}}{\sigma} \right] \text{ cu } \overline{X} : N\left[m_0, \frac{\sigma}{\sqrt{n}} \right] \quad (5.117)$$

Se remarcă că valoarea pragului de decizie λ este independentă de valoarea lui m în ipoteza $\mathbf{H_1}$. Urmează că testul este uniform cel mai puternic.

Variabila $Y = \frac{\overline{X} - m_0}{\sigma / \sqrt{n}}$ urmează o lege normală (într-adevăr σ este cunoscută şi joacă deci rolul unei constante) centrată şi redusă. Valoarea pragului va fi deci dedusă dintr-o tabelă a legii normale. La fel pentru eroarea de speța II-a şi pentru puterea testului.

Test bilateral. Fie **X** o variabilă aleatoare normală de medie **m** şi varianță σ^2 cunoscută. În vederea unui eşantion de n realizări independente $\mathbf{x_i}$ vrem să alegem între două ipoteze.

$$\begin{cases} H_0 : m = m_0 \\ H_1 : m \neq m_0 \end{cases} \quad (5.118)$$

Ca totdeauna, eroarea de speța I-a este fixată. Media m va fi estimată prin media aritmetică $\overline{x}$. Construcția testului este obținută observând că ipoteza $\mathbf{H_1}$ se poate să se descompune în două ipoteze elementare:

$$\begin{cases} H_1' : m < m_0 \\ H_1'' : m > m_0 \end{cases} \quad (5.119)$$

Fiecăreia dintre aceste două ipoteze i va fi asociat un prag de decizie λ' şi λ'' Se poate concluziona că testul nu va fi uniform puternic pentru că pragul de decizie λ depinde de sensul inegalității. Determinarea pragurilor este simplă pentru că cele două ipoteze sunt disjuncte. Avem:

$$\alpha = P[(\overline{X} \ge \lambda'') \; sau \; (\overline{X} \le \lambda')] = P(\overline{X} \ge \lambda'') + P(\overline{X} \le \lambda') = \alpha'' + \alpha' \quad (5.120)$$

Rezultă de aici o infinitate de valori posibile pentru λ' şi λ''. Totuşi, legea lui $\overline{X}$ fiind simetrică (lege normală), se ia în general

Material cu drept de autor

$\alpha'' + \alpha' = \alpha / 2$ ceea ce conduce la valori ale lui λ simetrice în raport cu $\mathbf{m_0}$. Fiecare caz de fapt este o aplicație a testului precedent dar pentru o valoare mai mică a lui α.

$$\frac{\alpha}{2} = P\left[Y = \frac{(\overline{X} - m_0).\sqrt{n}}{\sigma} \geq \frac{(\lambda - m_0).\sqrt{n}}{\sigma} \right] \quad \text{cu} \quad \overline{X} : N[m_0, \sigma] \; Y : N[0,1] \qquad (5.121)$$

Valoarea pragului este deci dedusă dintr-o tabelă a legii normale. La fel pentru eroarea de speța II-a şi pentru puterea testului.

B. Testul unei medii a legii normale, cu abaterea tip necunoscută

Cele două teste, bilateral şi unilateral, se construiesc după acelaşi procedeu. Valorile de decizie vor fi citite în tabele **t** Student cu **n-1** grade de libertate.

C. Testul unei varianțe a legii normale, media fiind cunoscută

Cele două teste, bilateral şi unilateral, se construiesc după acelaşi procedeu. Valorile de decizie vor fi citite în tabele de χ^2 cu **n** grade de libertate.

D. Testul unei varianțe al legii normale, media fiind necunoscută

Cele două teste, bilateral şi unilateral, se construiesc după acelaşi procedeu. Valorile de decizie vor fi citite în tabele de χ^2 cu **n-1** grade de libertate.

E. Testul unei proporții

Cele două teste, bilateral şi unilateral, se construiesc după acelaşi procedeu. Valorile de decizie vor fi citite în tabele a legii normale.

În cazul unui test bilateral, ne sprijinim pe faptul că proporția empirică $\mathbf{F_n}$ urmează aproximativ o lege normală de medie **p**, proporția teoretică, şi de abatere tip $\sqrt{[p.(1-p)]/n}$. Regiunea critică a testului este atunci:

$$|F_n - p| > u_{\alpha/2} \sqrt{\frac{p.(1-p)}{n}} \qquad (5.122)$$

unde $\mathbf{u_{\alpha/2}}$ este citită într-o tabelă a legii normale **N(0,1)**.

Exemplu de calcul. Dintr-un unui eşantion de **200** indivizi dintr-o comună, **30%** sunt favorabili unui centru comercial. Aceasta contrazice ipoteza după care un locuitor din **3** este favorabil ?

Material cu drept de autor

Rezolvare. Acest enunț conduce la construcția unui test bilateral de ipoteze de proporție:

$$\begin{cases} H_0 : p = 0{,}33 \\ H_1 : p \neq 0{,}33 \end{cases}$$

Cu α **= 0,05**, se citeşte **u = 1,96** de unde domeniul de acceptare:

$$\left|F_n - 0{,}33\right| > 1{,}96\sqrt{\frac{0{,}33 \times 0{,}67}{200}} = 0{,}065$$

Fie $\overline{W} = [0{,}265, 0{,}395]$. Cum $\left|f_n - 0{,}33\right| = 0{,}03 < 0{,}065$ nu se poate exclude H_0 cu pragul α **= 0,05**.

5.10. Teste de comparație

Fie X_1 şi X_2 două variabile aleatoare definite în două populații mamă comparabile (eventual egale). Legea lui X_1 (respectiv X_2) depinde de un parametru necunoscut θ_1 (respectiv θ_2). Dorim să testăm ipoteza dacă „aceşti doi parametri sunt egali" contra ipoteză complementară (cei doi parametri sunt diferiți):

$$H_0: \theta_1 = \theta_2 \text{ contra } H_1: \theta_1 \neq \theta_2. \qquad (5.123)$$

Pentru a efectua acest test, dispunem de un eşantion de volum n_1 (respectiv n_2) a variabilei X_1 (respectiv X_2) care permite o estimare punctuală T_{n1} (respectiv T_{n2}) a lui θ_1 (respectiv θ_2). Se presupune în plus că variabilele aleatoare X_1 şi X_2 sunt normale sau aproximativ normale.

Presupunând H_0 adevărată, se determină un risc de speța I-a α, o zonă de excludere asociată la două valori critice c_1 şi c_2, astfel că:

$$P(Z < c_1) = P(Z > c_2) = \alpha / 2 \qquad (5.124)$$

unde **Z** este o funcție de estimările T_{n1} (respectiv T_{n2}).

Dacă **Z** aparține zonei de excludere, se exclude H_0; dacă nu, se acceptă H_0 cu riscul α.

5.10.1. Comparația a două medii

Fie X_1 şi X_2 două variabile aleatoare cu legi normale de distribuție, mediile μ_1 şi μ_2, respectiv abaterile standard σ_1 şi σ_2. Se testează:

$$H_0: \mu_1 = \mu_2 \text{ contra } H_1: \mu_1 \neq \mu_2 \text{ cu riscul } \alpha. \qquad (5.125)$$

Se utilizează testul **t** Student (în versiunea sa cea mai generală).

Dispunem de două eşantioane de volume n_1 şi n_2, în care putem face estimările mediilor m_1 şi m_2, respectiv abaterile standard s_1 şi s_2.

Material cu drept de autor

Dacă abaterile standard σ_1 şi σ_2 sunt cunoscute, se calculează:

$$z = \frac{m_1 - m_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \qquad (5.126)$$

Se exclude H_0 cu risc α dacă $z \notin [-t_{1-\alpha/2}, t_{1-\alpha/2}]$ unde valoarea $t_{1-\alpha/2}$ este citită în tabela legii normale centrată redusă.

Dacă abaterile tip σ_1 şi σ_2 sunt necunoscute, trebuie să ţinem cont de volumul eşantioanelor.

a. dacă n_1 şi n_2 sunt amândouă superioare lui 30, se calculează:

$$z = \frac{m_1 - m_2}{\sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}} \qquad (5.127)$$

Se exclude H_0 cu risc α dacă $z \notin [-t_{1-\alpha/2}, t_{1-\alpha/2}]$ unde valoarea $t_{1-\alpha/2}$ este citită în tabela legii normale centrată redusă.

b. dacă n_1 sau n_2 este inferior lui 30 şi $\sigma_1 = \sigma_2$ se calculează:

$$z = \frac{m_1 - m_2}{\hat{\sigma}.\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \text{ unde } \hat{\sigma} = \sqrt{\frac{n_1.s_1^2 + n_2.s_2^2}{n_1 + n_2 - 2}} \qquad (5.128)$$

Se exclude H_0 cu risc α dacă

$$z \notin [-t_{1-\alpha/2;n_1+n_2-2}, t_{1-\alpha/2;n_1+n_2-2}]$$

unde valoarea $t_{1-\alpha/2;n_1+n_2-2}$ este citită în tabela **t** a lui Student la $n_1 + n_2 - 2$ grade de libertate.

c. Dacă n_1 sau n_2 este inferior lui 30 şi $\sigma_1 \neq \sigma_2$ se calculează:

$$z = \frac{m_1 - m_2}{\sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}} \qquad (5.129)$$

Se exclude H_0 cu risc α dacă $z \notin [-t_{1-\alpha/2;\upsilon}, t_{1-\alpha/2;\upsilon}]$ unde valoarea $t_{1-\alpha/2;\upsilon}$ este citită în tabela **t** a lui Student cu υ grade de libertate; υ este întregul cel mai apropiat de:

Material cu drept de autor

$$\frac{\left[\frac{s_1^2}{n_1-1}+\frac{s_2^2}{n_2-1}\right]}{\frac{s_1^4}{(n_1-1).n_1^2}+\frac{s_2^4}{(n_2-1).n_2^2}} \qquad (5.130)$$

Testul Student este destul de robust dar dacă ne depărtăm prea mult de condițiile de normalitate este preferabilă utilizarea unui test neparametric.

5.10.2. Comparația a două varianțe

Cu acelaşi notații ca în testul precedent se testează

H_0: $\sigma_1 = \sigma_2$ contra H_1: $\sigma_1 \neq \sigma_2$ cu riscul α. (5.131)

Se calculează:

$$\hat{s}_1^2 = \frac{n_1.s_1^2}{n_1-1};\ \hat{s}_2^2 = \frac{n_2.s_2^2}{n_2-1} \text{ şi } z = \frac{\hat{s}_1^2}{\hat{s}_2^2} \qquad (5.132)$$

Se exclude H_0 cu risc α dacă:

$$z \notin [F_{\alpha/2(n_1-1,n_2-1)}, F_{1-\alpha/2(n_1-1,n_2-1)}]$$

unde valoarea F_α este citită în tabela Fisher-Snedecor n_1-1 şi n_2-1 reprezintă numărul gradelor de libertate.

Observație:

$$F_{\alpha/2(n_1-1,n_2-1)} = \frac{1}{F_{1-\alpha/2(n_1-1,n_2-1)}} \qquad (5.133)$$

5.10.3. Comparația a două proporții

Fie p_1 (respectiv p_2) proporția de indivizi de o anumită modalitate **A** în populația mamă M_1 (respectiv M_2). Se extrage un eşantion de volum n_1 (respectiv n_2) din populația M_1 (respectiv M_2). Se testează plecând de la aceste eşantioane. Se dispune de o estimare: f_1 (respectiv f_2) de p_1 (respectiv p_2) care urmează o lege statistică F_1 (respectiv F_2).

H_0: $p_1 = p_2$ contra H_1: $p_1 \neq p_2$ cu riscul α. (5.134)

Se presupune că $n_1.F_1$ şi $n_2.F_2$ urmează aproximativ legi normale. Se calculează:

Material cu drept de autor

$$\hat{p} = \frac{n_1.f_1 + n_2.f_2}{n_1 + n_2} \text{ apoi } z = \frac{f_1 - f_2}{\sqrt{\hat{p}.(1-\hat{p}).\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad (5.135)$$

Se exclude $\mathbf{H_0}$ cu risc α dacă $\mathbf{z \notin [-t_{1-\alpha/2}, t_{1-\alpha/2}]}$ unde valoarea $\mathbf{t_{1-\alpha/2}}$ este citită în tabela legii normale centrate redusă.

5.11. Test de adecvare

În această parte, se presupune că legea de probabilitate a variabilei aleatoare **X**, din care avem un eşantion, este necunoscută. O primă remarcă se impune. Testele de adecvare nu permit să se găsească legea unei variabilă aleatoare, ci numai de a accepta sau exclude o ipoteză simplă emisă apriori.

Astfel, este necesar să se facă un studiu sumar prealabil al eşantionului în scopul de a formula ipoteze plauzibile privitoare la legea de probabilitate a lui **X**: la variabila aleatoare **X** este discretă sau continuă ? este definită pentru orice **x**, sau numai pentru **x > 0** ? histograma în frecvență obținută este simetrică în raport cu valoarea medie ? există vreo relație simplă între media estimată şi varianța estimată ? Răspunsurile la aceste diferite întrebări, la fel ca şi natura variabilei reprezentate prin **X** permit în cele mai multe cazuri să se emită o ipoteză plauzibilă.

5.12. Testul χ^2

Fie $\{\mathbf{x_1, x_2, ..., x_n}\}$ un eşantion de **n** realizări independente ale variabilei aleatoare **X**. Fie **L(x)** legea de distribuție necunoscută a lui **X**. Ipoteza de plecare va fi că legea de distribuție **L*(x)**. Aceasta permite să formulăm testul:

$$H_0 : L(x) = L^*(x)$$
$$H_1 : L(x) \neq L^*(x) \qquad (5.136)$$

Parametrii lui **L*(x)** vor fi cunoscuți, fie estimați.

Plecând de la eşantion se construieşte o histogramă în frecvență de **k** clase $\mathbf{C_i}$. Se notează $\mathbf{O_i}$ numărul de observații a lui **X** făcute în clasa $\mathbf{C_i}$ (bineînțeles $\Sigma_i \mathbf{O_i} = \mathbf{n}$). Dacă variabila aleatoare urmează legea **L*(x)** atunci efectivul teoretic $\mathbf{E_i}$ din clasa $\mathbf{C_i}$ este dat de: $\mathbf{E_i = n.p^*_i}$ unde $\mathbf{p^*_i}$ este probabilitatea ca variabila aleatoare **X** ce urmează legea **L*** să ia o valoare în domeniul ce defineşte clasa $\mathbf{C_i}$.

Abaterea între realitatea rezultată din eşantion şi teoria rezultată din ipoteza $\mathbf{H_0}$ este măsurată prin indicatorul:

Material cu drept de autor

$$I = \sum_{i=1}^{k} \frac{(n.p_i^* - O_i)^2}{n.p_i^*} \qquad (5.137)$$

În ipoteza $\mathbf{H_0}$ se poate considera ca abaterea $\mathbf{E_i}$ - $\mathbf{O_i}$, între distribuția teoretică şi distribuția empirică este distribuită normal. În aceste condiții, **I** tinde către o lege χ^2, cu υ grade de libertate (υ este egal cu numărul de clase – 1 – numărul de parametri necesar specificării complete a lui $\mathbf{p^*_i}$).

Regiunea de acceptare a testului este intervalul $(0, \chi^2_{\nu,1-\alpha})$ astfel că probabilitatea ca o variabilă χ^2 cu υ grade de libertate să ia o valoare în acest interval să fie egală cu **1**-α (α fiind eroarea de speța I-a relativă la test). Dacă valoarea indicatorului este superioară lui $\chi^2_{\nu,1-\alpha}$ atunci se decide ipoteza $\mathbf{H_1}$.

Nu este deloc posibil să se determine eroarea de speța II-a (şi deci puterea testului), legea de probabilitate a lui **X** nefiind specificată în ipoteza $\mathbf{H_1}$.

Exeplul de calcul 5.12. Managerul unui restaurant doreşte să determine dacă o campanie publicitară a mărit semnificativ media încasărilor zilnice. El culege date privitoare la 50 de zile înaintea campaniei şi la 30 de zile după încheierea campaniei publicitare. Rezultatele pentru cele două perioade sunt:

Înaintea campaniei	**După campanie**
$n_1 = 50$	$n_2 = 50$
$\bar{x}_1 = 12{,}55 \ \ u.m.$	$\bar{x}_2 = 13{,}30 \ \ u.m.$
$\sigma_1 = 2{,}15 \ \ u.m.$	$\sigma_2 = 2{,}38 \ \ u.m.$

Informațiile obținute sunt suficiente pentru a susține ipoteza, conform căreia mediile încasărilor diferă semnificativ, adică sunt semnificativ mai mari după campania publicitară ? Să se utilizeze o probabilitate de 95%.

Rezolvare.

$\mathbf{H_0}$: $\bar{x}_{01} - \bar{x}_{02} = 0 \ u.m.$;

$\mathbf{H_a}$: $\bar{x}_{01} - \bar{x}_{02} < 0 \ u.m. \ (\bar{x}_{01} < \bar{x}_{02})$

Testul statistic:

$$z = \frac{(\bar{x}_{01} - \bar{x}_{02}) - 0}{\sigma_{(\bar{x}_1 - \bar{x}_2)}}$$

Presupunând că cele două eşantioane sunt independente:

$$\sigma_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Material cu drept de autor

$$z = \frac{12,55 - 13,30}{\sqrt{\frac{(2,15)^2}{50} + \frac{(2,38)^2}{50}}} = \frac{12,55 - 13,30}{0,53} = -1,42$$

Domeniul de respingere: $|z| > |z_{0.05}| = 1,645$.

Cum $|z_{calc}| < |z_{\alpha=0.05}|$, rezultă că nu sunt motive a se respinge ipoteza nulă, conform căreia diferența dintre cele două medii nu este semnificativ diferită de **0** (nu sunt suficiente dovezi pentru a susține creşterea încasărilor după campania publicitară).

5.13. Serii de distribuție bidimensionale

A. Calcule cu frecvențe absolute. O serie de distribuție bidimensională se prezintă în tabelul următor:

Valorile caracteristicii de grupare **X**	Variantele sau valorile caracteristicii dependente **Y**						Volumul grupei (n_i)	Medii pe grupe ($\bar{y}$)
	y_1	y_2	...	y_j	...	y_m		
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1m}	$n_{1\bullet}$	$\bar{y}_1$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2m}	$n_{2\bullet}$	$\bar{y}_2$
...	...	...	...	...	...	...	...	...
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{im}	$n_{i\bullet}$	$\bar{y}_i$
...	...	...	...	...	...	...	...	...
x_r	n_{r1}	n_{r2}	...	n_{rj}	...	n_{rm}	$n_{r\bullet}$	$\bar{y}_r$
Total	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet j}$	...	$n_{\bullet m}$	$\Sigma n_{i\bullet} = \Sigma n_{\bullet j}$	$\bar{\bar{y}}$

Volumul (frecvența) grupei i:

$$\sum_{j=1}^{m} n_{ij} = n_{i\bullet}$$

Mărimi medii.

- **Mediile de grupă** ($\bar{y}_i$):

Material cu drept de autor

$$\bar{y}_i = \frac{\sum_{j=1}^{m} y_j n_{ij}}{\sum_{j=1}^{m} n_{ij}}$$

- **Media pe total**:

$$\bar{\bar{y}} = \frac{\sum_{j=1}^{m} y_j n_{\bullet j}}{\sum_{j=1}^{m} n_{\bullet j}} \quad sau \quad \bar{\bar{y}} = \frac{\sum_{i=1}^{r} \bar{y}_i n_{i\bullet}}{\sum_{i=1}^{r} n_{i\bullet}}$$

- Indicatorii variației
- Dispersia de grupă sau dispersia parțială σ_i^2:

$$\sigma_i^2 = \frac{\sum_{i=1}^{m} (y_j - \bar{y}_i)^2 . n_{ij}}{\sum_{j=1}^{m} n_{ij}}$$

unde $\mathbf{y_j}$ reprezintă variabila sau mijlocul intervalului **j** al caracteristicii dependente; $\bar{y}_i$ media grupei **i**; $\mathbf{n_{ij}}$ – frecvențele corespunzătoare fiecărei variabile (interval de valori) din cadrul grupei.

- **Media dispersiilor de grupă** ($\bar{\sigma}^2 = \sigma_{y/r}^2$):

$$\sigma_{y/r}^2 = \bar{\sigma}^2 = \frac{\sum_{i=1}^{r} \sigma_i^2 n_{i\bullet}}{\sum_{i=1}^{r} n_{i\bullet}}$$

unde: σ_i^2 - dispersia grupei i; $n_{i\bullet}$ - volumul grupei i.

- **Dispersia dintre grupe** $\delta^2 = \sigma_{y/x}^2$:

$$\sigma_{y/x}^2 = \delta^2 = \frac{\sum_{i=1}^{r} (\bar{y}_i - \bar{\bar{y}})^2 . n_{i\bullet}}{\sum_{i=1}^{r} n_{i\bullet}}$$

- **Dispersia totală** $\sigma^2 = \sigma_y^2$:

Material cu drept de autor

$$\sigma^2 = \sigma_y^2 = \frac{\sum_{j=1}^{m} (y_j - \bar{\bar{y}})^2 n_{\bullet j}}{\sum_{j=1}^{m} n_{\bullet j}}$$

- **Regula adunării dispersiilor**:

$$\sigma^2 = \sigma_{y/x}^2 + \sigma_{y/r}^2$$

Pe baza regulii de adunare a dispersiilor se pot calcula indicatorii statistici cu caracter de mărimi relative de structură:

- **Gradul de determinație** $R_{y/x}^2$:

$$R_{y/x}^2 = \frac{\sigma_{y/x}^2}{\sigma^2} \times 100$$

Dacă $R_{y/x}^2$>50% se admite că factorul de grupare este hotărâtor (semnificativ, determinant) pentru variația factorului determinant **Y**.

- **Gradul de nedeterminație** $K_{y/x}^2$:

$$K_{y/x}^2 = \frac{\sigma_{y/r}^2}{\sigma^2} \times 100$$

Abaterea medie pătratică la nivelul grupei: $\sigma_i = \sqrt{\sigma_i^2}$

Abaterea medie pătratică la nivelul grupei: $\sigma = \sqrt{\sigma^2}$

Coeficientul de variație la nivelul grupei: $\nu_i = (\sigma_i / \bar{y}_i) \times 100$

unde σ_i – abaterea medie pătratică a grupei **i**; $\bar{y}_i$ - media grupei **i**.

Coeficientul de variație la nivelul grupei: $\nu = (\sigma / \bar{\bar{y}}) \times 100$

unde: σ - abaterea medie pătratică pe total; $\bar{\bar{y}}$ - media pe total.

B. Calculul cu frecvențe relative. O serie bidimensională cu frecvențe relative se prezintă ca în tabelul următor:

Valorile caracteristicii de grupare **X**	Variantele sau valorile caracteristicii dependente **Y**						Total (%)	Ponderea grupei $n_i\%$
	y_1	y_2	...	y_j	...	y_m		
x_1	n_{11}^*	n_{12}^*	...	n_{1j}^*	...	n_{1m}^*	**100**	$n_1\%$
x_2	n_{21}^*	n_{22}^*	...	n_{2j}^*	...	n_{2m}^*	**100**	$n_2\%$

Material cu drept de autor

...	...	...	...	...	...	...	...	
x_i	n_{i1}^*	n_{i2}^*	...	n_{ij}^*	...	n_{im}^*	100	$n_i\%$
...	...	...	...	...	...	...	...	
x_r	n_{r1}^*	n_{r2}^*	...	n_{rj}^*	...	$n_{r\,m}^*$	100	$n_r\%$
Total								100

- **Mediile de grupă** $\overline{y}_i$:

$$\overline{y}_i = \frac{\sum_{j=1}^{m} y_j . n_{i\,j}^*}{100}$$

- **Media pe total** $\overline{\overline{y}}$:

$$\overline{\overline{y}} = \frac{\sum_{i=1}^{r} \overline{y}_i n_{i(\%)}}{100}$$

- **Dispersia de grupă sau dispersia parțială** σ_i^2 :

$$\sigma_i^2 = \frac{\sum_{i=1}^{m} (y_j - \overline{y}_i)^2 n_{i\,j}^*}{100}$$

- **Media dispersiilor de grupă** $\overline{\sigma}^2 = \sigma_{y/r}^2$:

$$\overline{\sigma}^2 = \frac{\sum_{i=1}^{r} \sigma_i^2 n_{i(\%)}}{100} :$$

- **Dispersia dintre grupe** $\delta^2 = \sigma_{y/x}^2$:

$$\delta^2 = \sigma_{y/x}^2 = \frac{\sum_{i=1}^{r} (\overline{y}_i - \overline{\overline{y}})^2 . n_{i(\%)}}{100}$$

- **Dispersia totală** $\sigma^2 = \sigma_y^2$:

$$\sigma^2 = \sigma_y^2 = \overline{\sigma}^2 + \delta^2$$

Material cu drept de autor

C. Calculul cu frecvențe alternative.

- **Dispersia de grupă sau dispersia parțială** $\sigma_{p_i}^2$:

$$\sigma_{p_i}^2 = p_i.q_i \quad sau \quad \sigma_{p_i}^2 = p_i.(1-p_i)$$

în care: p_i – reprezintă medii de grupă: q_i – frecvențele relative ale unităților care nu posedă caracteristica în fiecare grupă.

Media dispersiilor parțiale $\overline{\sigma}_p^2$:

$$\overline{\sigma}_p^2 = \frac{\sum_{i=1}^{r} \sigma_{p_i}^2 . N_i}{\sum_{i=1}^{r} N_i}$$

în care N_i reprezintă numărul total al unităților observate în fiecare grupă.

- **Dispersia dintre grupe** δ_p^2 :

$$\delta_p^2 = \frac{\sum_{i=1}^{r} (p_i - \overline{p})^2 . N_i}{\sum_{i=1}^{r} N_i}$$

în care $\overline{p}$ **este media caracteristicii alternative pe întreaga colectivitate.**

- **Dispersia totală** σ_p^2 :

$$\sigma_p^2 = p.q$$

- **Regula adunării dispersiilor**:

$$\sigma_p^2 = \overline{\sigma}_p^2 + \delta_p^2$$

Verificarea semnificației factorului de grupare folosind testul „F":

$$F_{calculat} = \frac{S_{y/x}^2}{S_{y/r}^2}$$

unde:

$$S_{y/x}^2 = \frac{\sum_{i=1}^{r} (\overline{y}_i - \overline{y})^2 . n_i}{r-1}; \quad S_{y/r}^2 = \frac{\sum_{i=1}^{r} \sum_{j=1}^{m} (y_j - \overline{y}_i)^2 . n_{ij}}{n-r};$$

Dacă $F_{calculat} > F_{tabelat}$ factorul de grupare este semnificativ.
Dacă $F_{calculat} < F_{tabelat}$ factorul de grupare nu este semnificativ.

Material cu drept de autor

$F_{tabelat}$ se determină în funcție de un anumit nivel de semnificație (de exemplu 0,05) şi de numărul gradelor de libertate $f_1 = r - 1$ şi $f_2 = n - r$.

Exemplu de calcul 5.13. Se cunosc următoarele date privind vânzările dintr-un complex comercial în funcție de vechimea în muncă a vânzătorilor şi valoarea vânzărilor realizate într-o săptămână:

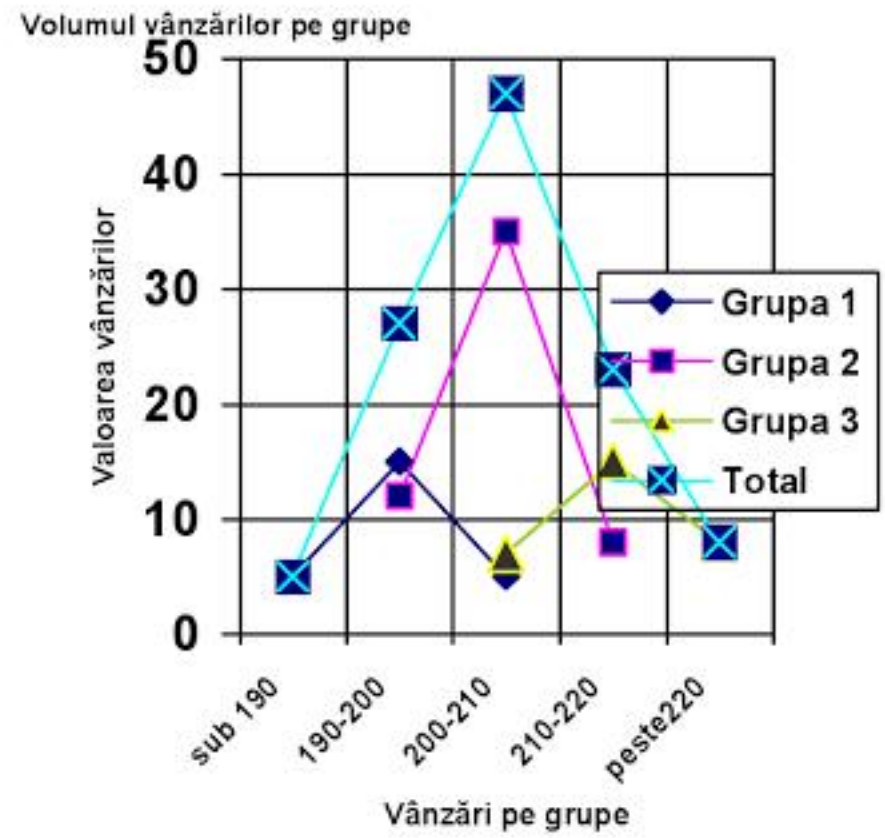
Grupe de vânzători după vechime, ani	Subgrupe de vânzători după volumul vânzărilor (u.m.)					Total
	<190	**190-200**	**200-210**	**210-220**	**>220**	
Sub 10	5	15	5	-	-	25
10 – 20	-	12	35	8	-	55
20 şi peste	-	-	7	15	8	30
Total	5	27	47	23	8	110

Se cere:

1. Poligonul frecvențelor privind repartiția vânzătorilor după volumul vânzărilor pe total şi pe grupe de vechime;
2. Calculul mediilor pe grupe de vechime şi pe total;
3. Indicatorii sintetici ai variației pe fiecare grupă şi pe total;
4. Interpretarea gradului de omogenitate pe grupe şi pe total;
5. Verificarea regulii de adunare a dispersiilor;
6. Ce indicatori sintetici se pot calcula pe baza regulii de adunare a dispersiilor şi cum se interpretează statistic aceşti indicatori;
7. Calculul şi interpretarea dispersiilor pentru caracteristica „vânzători care se află peste media vânzărilor pe total".

Rezolvare.

1. Poligonul frecvențelor privind repartiția vânzătorilor după volumul vânzărilor pe total şi pe grupe de vechime:



Material cu drept de autor

2. Calculul mediilor pe grupe de vechime şi pe total:

• **Mediile pe grupe** ($\bar{y}_i$):

$$\bar{y}_i = \frac{\sum_{j=1}^{m} y_j n_{ij}}{\sum_{j=1}^{m} n_{ij}} :$$

$$\bar{y}_1 = \frac{185.5 + 195.15 + 205.5}{25} = 195\ u.m./vanzator$$

$$\bar{y}_2 = \frac{195.12 + 205.35 + 215.8}{55} = 204{,}27\ u.m./vanzator$$

$$\bar{y}_3 = \frac{205.7 + 215.15 + 225.8}{30} = 215{,}33\ u.m./vanzator$$

Media generală $\bar{y}$:

- independent:

$$\bar{y} = \frac{\sum_{j=1}^{m} y_j . n_{\bullet j}}{\sum_{j=1}^{m} n_{\bullet j}}$$

$$\bar{y} = \frac{185.5 + 195.27 + 205.47 + 215.23 + 225.8}{110} = 205{,}18 u.m./vanzator$$

- pe baza mediilor de grupă:

$$\bar{y} = \frac{\sum_{i=1}^{r} \bar{y}_i . n_{i\bullet}}{\sum_{i=1}^{r} n_{i\bullet}}$$

$$\bar{y} = \frac{195.25 + 204{,}27.55 + 215{,}33.30}{110} = 205{,}18\ u.m./vanyator$$

3. Calculul indicatorilor sintetici ai variației:

Material cu drept de autor

$$\sigma_i^2 = \frac{\sum_{j=1}^{m}(y_j - \bar{y}_i)^2 . n_{ij}}{\sum_{j=1}^{m} n_{ij}}$$

- **Dispersiile de grupă** σ_i^2:

$$\sigma_1^2 = \frac{(185-195)^2.5 + (195-195)^2.15 + (205-195)^2.5}{25} = 40$$

$$\sigma_2^2 = \frac{(195-204{,}27)^2.12 + (205-204{,}27)^2.35 + (215-204{,}27)^2.8}{55} = 35{,}83$$

$$\sigma_3^2 = \frac{(205-215{,}33)^2.7 + (215-215{,}33)^2.15 + (225-215{,}33)^2.8}{30} = 49{,}89$$

- **Media dispersiilor parțiale** $\overline{\sigma}^2$:

$$\overline{\sigma}^2 = \frac{\sum_{i=1}^{r}\sigma_i^2 . n_{i\bullet}}{\sum_{i=1}^{r} n_{i\bullet}} = \frac{40.25 + 35{,}83.55 + 49{,}89.30}{110} = 40{,}63$$

- **Dispersia dintre grupe** $\delta^2 = \sigma_{y/x}^2$:

$$\delta^2 = \frac{\sum_{i=1}^{r}(\bar{y}_i - \bar{y})^2 . n_{i\bullet}}{\sum_{i=1}^{r} n_{i\bullet}} =$$

$$= \frac{(195-205{,}18)^2.25 + (204{,}27-205{,}18)^2.55 + (215{,}33-205{,}18)^2.30}{110} = 52{,}06$$

- **Dispersia totală** σ^2:

$$\sigma^2 = \frac{\sum_{j=1}^{m}(y_j - \bar{y})^2 n_{\bullet j}}{\sum_{j=1}^{m} n_{\bullet j}} = \frac{(185-205{,}18)^2.5 + (195-205{,}18)^2.27}{110} +$$

$$+ \frac{(205-205{,}18)^2.47 + (215-205{,}18)^2.23 + (225-205{,}18)^2.8}{110} = 92{,}69$$

Material cu drept de autor

4. Aprecierea gradului de omogenitate pe grupe şi pe total: Calculul coeficienților de variație:

• **Pe grupe**:

$$\sigma_1 = \sqrt{\sigma_1^2} = \sqrt{40} = 6{,}32\ u.m./vanzator$$

$$\nu_1 = \frac{\sigma_1}{\bar{y}_1} \times 100 = \frac{6{,}32}{195} \times 100 = 3{,}24\%$$

$$\sigma_2 = \sqrt{\sigma_2^2} = \sqrt{35{,}83} = 5{,}98\ u.m./vanzator$$

$$\nu_2 = \frac{\sigma_2}{\bar{y}_2} \times 100 = \frac{5{,}98}{204{,}27} \times 100 = 2{,}93\%$$

$$\sigma_3 = \sqrt{\sigma_3^2} = \sqrt{49{,}89} = 7{,}06\ u.m./vanzator$$

$$\nu_3 = \frac{\sigma_3}{\bar{y}_3} \times 100 = \frac{7{,}06}{215{,}33} \times 100 = 3{,}18\%$$

• **Pe total**:

$$\sigma = \sqrt{\sigma^2} = \sqrt{92{,}69} = 9{,}62\ u.m./vanzator$$

$$\nu = \frac{\sigma}{\bar{y}} \times 100 = \frac{9{,}62}{205{,}18} \times 100 = 4{,}69\%$$

Comparând rezultatele se constată că:

➲ fiecare grupă luată separat este mai omogenă decât colectivitatea generală din care a fost extrasă;

➲ grupa doua este mai omogenă decât celelalte două;

➲ valorile mici ale coeficienților de variație calculați pe fiecare grupă şi pe total atestă un grad de omogenitate ridicat al grupelor şi colectivității totale şi deci un grad de reprezentativitate corespunzător pentru mediile care le caracterizează.

5. Verificarea regulii de adunare a dispersiilor:

$$\sigma^2 = \delta^2 + \bar{\sigma}^2 \Rightarrow 92{,}69 = 52{,}06 + 40{,}63$$

6. Pe baza regulii de adunare a dispersiilor se pot calcula alți doi indicatori statistici:

• Gradul de determinație $R^2_{y/x}$:

$$R^2_{y/x} = \frac{\delta^2}{\sigma^2} \times 100 = \frac{52{,}06}{92{,}69} \times 100 = 56{,}16\%$$

• Gradul de nedeterminație $K^2_{y/x}$:

Material cu drept de autor

$$K^2_{y/x} = \frac{\overline{\sigma}^2}{\sigma^2} \times 100 = \frac{40,63}{92,69} \times 100 = 43,84\%$$

Se poate afirma că 56,16% din variația totală a volumului vânzărilor este explicată prin variația produsă de factorul de grupare (vechimea – factor determinant întrucât $R^2_{y/x}$>50%), restul de 43,84% fiind influența relativă a celorlalți factori neînregistrați.

7. Calculul şi interpretarea dispersiilor pentru „vânzătorii care se află peste volumul mediu al vânzărilor pe total":

☞ **Calculul mediilor**:

• mediile de grupă:

$$w_i = \frac{m_i}{n_i} \Rightarrow w_1 = \frac{0}{25} = 0;\ w_2 = \frac{8}{55} = 0,1455;$$

$$w_3 = \frac{23}{30} = 0,7667$$

• media pe total: $w = \frac{m}{n} = \frac{31}{110} = 0,2818$

☞ **Calculul dispersiilor**:

• **dispersiile de grupă**: $\sigma^2_{w_i} = w_i.(1-w_i)$, *de unde* :

$$\sigma^2_{w_1} = w_1.(1-w_1) = 0.(1-0) = 0$$

$$\sigma^2_{w_2} = w_2.(1-w_2) = 0,1455.(1-0,1455) = 0,1243$$

$$\sigma^2_{w_3} = w_3.(1-w_3) = 0,7667.(1-0,7667) = 0,1789$$

• **Media dispersiilor de grupă** $\overline{\sigma}^2_w$:

$$\overline{\sigma}^2_w = \frac{\sum_{i=1}^{r} \sigma^2_{w_i}.n_i}{\sum_{i=1}^{r} n_i} =$$

$$= \frac{0.25 + 0,1243.55 + 0,1789.30}{110} = 0,1109$$

• **Dispersia dintre grupe** δ^2_w :

Material cu drept de autor

$$\delta_w^2 = \frac{\sum_{i=1}^{r}(w_i - w)^2 . n_i}{\sum_{i=1}^{r} n_i} = \frac{(0-0,2818)^2 .25 + (0,1455-0,2818)^2 .55}{110} +$$

$$+ \frac{(0,7667-0,2818)^2}{110} = 0,0914$$

- **Dispersia totală** σ_w^2:

$$\sigma_w^2 = w.(1-w) = 0,2818.(1-0,2818) = 0,2023$$

Regula adunării dispersiilor se păstrează şi în cazul caracteristicii alternative:

$$\sigma_w^2 = \overline{\sigma}_w^2 + \delta_w^2 \quad \Rightarrow \quad 0,2023 = 0,1109 + 0,0914$$

Material cu drept de autor

Cap.6. ANALIZA DISPERSIONALĂ

6.1. Consideraţii introductive

Analiza dispersională, numită şi analiza variaţei, a fost introdusă în calculele statistice de către matematicianul englez R. A. Fisher, în preocupările lui de a pune la punct o serie de principii ale planificării şi analizei experimentelor care au revoluţionat de atunci metodologia cercetării în agricultură, sub denumirea de *Analysis of Variance*, de unde şi denumirea întâlnită în manualele de specialitate de metoda ANOVA.

Contribuţiile lui R. A. Fisher în domeniul statisticii matematice sunt concretizate în două lucrări de bază în statistică, şi anume: *Statistical Methods for research design of Experiments* (Metode statistice pentru cercetători ştiinţifici), publicată în 1925 şi *The Design of Experiments* (Proiectarea experimentelor), publicată în 1935. În aceste lucrări se află descrise principiile filozofice şi tehnicile principiale ale domeniilor respective - ANOVA şi proiectarea experimentelor.

Statistica este justificată şi printr-o lege a naturii: "*variabilitatea în repetare şi nu reproducerea identică*". Indivizii unei specii se aseamănă dar nu sunt identici. De aceea, nici în activităţile umane nu se pot reproduce identic acţiunile întreprinse, cu rezultatele identice. Dacă se repetă de mai multe ori măsurarea unei caracteristici oarecare ce defineşte o situaţie sau un proces, rezultatele ce se vor obţine nu vor fi niciodată identice, ci vor prezenta o variabilitate mai mică sau mai mare.

Variabilitatea rezultatelor obţinute în practica activităţilor umane se poate datora unor factori cu efecte sistematice, apoi a unor factori aleatorii de fluctuaţie, inerenţi şi inevitabili, care definesc variabilitatea experimentală de fluctuaţie a cercetărilor experimentale pe populaţii statistice.

Analiza dispersională oferă posibilitatea de a diviza variabilitatea totală în: variabilitatea datorată factorilor cu efecte sistematice, plus variabilitatea datorată factorilor cu efecte aleatoare şi o variabilitate reziduală (diferenţa până la variabilitatea totală) care nu este în fond decât variabilitatea experimentală menţionată. **Pe baza acestei descompuneri se pot calcula dispersiile parţiale aferente diverşilor factori, după care semnificaţia lor se verifică cu ajutorul testului F a lui Fischer**.

Principial, datele de măsurare se grupează în raport de unul sau mai multe criterii, după care se scot în evidenţă efectele în funcţie de influenţa specifică a acestor criterii. odată efectele puse în evidenţă, testarea se face prin compararea dispersiilor produse de diferiţi factori a căror variaţie nu o cunoaştem şi avem să o descoperim, sau pe care o facem noi să varieze - cu dispersia produsă de factorii întâmplători care acţionează inevitabil asupra procesului (dispersia reziduală sau experimentală).

Material cu drept de autor

Aplicabilitatea analizei dispersionale este condiționată de: distribuția normală a datelor de observație (o abatere moderată poate fi acceptată); omogenitatea dispersiilor de selecție (dispersia experimentală comună), aditivitatea efectelor factorilor.

Numărul criteriilor după care se grupează datele depinde de numărul de factori luați în considerare. Dacă se ia în considerare un singur factor variabil analiza dispersională se numeşte unifactorială. Dacă se urmăreşte influența simultană a doi sau mai mulți factori analiza dispersională respectivă se numeşte bi- sau poli(multi)factorială.

Serviciile pe care analiza dispersională le oferă analizei statistice a fenomenelor tehnico-economice pot fi concretizate prin:

- oferă posibilitatea comparării mediilor rezultatelor mai multor analişti sau mai multor laboratoare, în vederea descoperirii unor eventuale erori sistematice,
- oferă posibilitatea descompunerii erorii totale a unei metode de analiză statistică în erorile parțiale ale fazelor metodei, relevând astfel fazele ce trebuiesc îmbunătățite;
- oferă posibilitatea de a stabili dacă un fenomen supus analizei statistice este omogen sau nu şi pe această bază dă posibilitatea calculării erorii de luare a probei;
- în cazul elaborării unei metode noi de cercetare a fenomenelor tehnico-economice oferă posibilitatea punerii în evidență a efectului factorilor implicați, a interacțiunii dintre factori şi a factorilor nesemnificativi;
- în cazul elaborării unei metode noi de analiză a factorilor de influență în combinare cu analiza de regresie, oferă posibilitatea optimizării experimentării.

La realizarea cercetărilor experimentale se obțin indicatori statistici: valori medii şi dispersii individuale, pe serii de date experimentale. Problema de bază care se pune este de a determina în ce măsură valorile medii şi dispersiile exprimă aceeaşi valoare medie şi dispersie a populației statistice. Pentru aceasta se utilizează diferite criterii statistice de evaluare, în anumite limite ale intervalului de încredere, egalitatea valorilor medii şi dispersiilor de selecție. Dacă pentru egalitatea dispersiilor au fost elaborate metode adecvate de evaluare a egalității dispersiilor de selecție, în sensul evaluării încrederii, cu o anumită probabilitate, că acestea exprimă aceeaşi dispersie generală a populației statistice, pentru egalitatea mediilor se apelează cu succes la analiza dispersională.

Din punct de vedere formal, ***analiza dispersională constituie instrumentul cel mai bun de verificare a ipotezei statistice a omogenității mediilor mai multor populații normale, în anumite condiții impuse acestor populații***.

Soluția - cea mai bună până acum - a problemei comparării mediilor "normale" a fost oferită de analiza dispersională. Ca subdomeniu al statisticii matematice, analiza dispersională a luat, în ultimul timp, un avânt deosebit, pe de o parte datorită avantajelor mari oferite la analiza omogenității mediilor

Material cu drept de autor

statistice, iar pe de altă parte datorită necesității optimizării cercetărilor experimentale pe baza metodei planificării experimentelor.

6.2. Criteriul de egalitate a două dispersii

Verificarea omogenității dispersiilor de selecție urmăreşte evidențierea faptului că toate dispersiile de selecție estimează dispersia comună a aceleiaşi populații statistice.

Dacă ipoteza omogenității dispersiilor de selecție se confirmă, înseamnă că dispersia generală σ^2 a populației statistice se poate estima pe baza mediei aritmetice ponderate a celor **m** dispersii de selecție. Dacă însă pe baza regulilor acestui criteriu ipoteza omogenității dispersiilor de selecție trebuie respinsă este necesar să se analizeze situația şi să se elimine dispersiile de selecție care nu se încadrează în criteriu omogenității acestora.

Se efectuează, de exemplu, două cercetări succesive, la un anumit interval de timp, ale unei caracteristici de calitate pentru un anumit produs. Se vor obține două valori pentru dispersiile celor două serii de cercetări statistice. Dacă cele două dispersii diferă între ele, se pune problema dacă exprimă aceeaşi calitate a caracteristicii urmărite. Comparația între cele două dispersii se face cu ajutorul criteriului **F** al lui Fischer.

Astfel, se calculează cele două dispersii obținându-se s_1^2 şi respectiv s_2^2. Se calculează apoi raportul:

$$F = \frac{s_1^2}{s_2^2} \text{ dacă } s_1^2 > s_2^2 \qquad (6.1)$$

respectiv:

$$F = \frac{s_2^2}{s_1^2} \text{ dacă } s_2^2 > s_1^2 \qquad (6.2)$$

Valoarea **F** obținută se compară cu valoarea F_P din anexa 5, pentru un anumit nivel de încredere **P** şi numărul gradelor de libertate ν_1 şi ν_2. Dacă $F \leq F_P$ se acceptă ipoteza egalității celor două dispersii, în sensul că exprimă aceeaşi dispersie generală σ^2.

Exemplul de calcul 6.1. Se supune analizei statistice o anumită caracteristică de calitate a unui produs, la două momente de realizare a lui diferite. La primul moment se iau în considerare 30 probe elementare şi după prelucrarea statistică a datelor înregistrate se obțin: $\bar{x}_1 = 40,1$ şi $s_1^2 = 0,82$. La al doilea moment se iau în considerare 20 de probe elementare şi după prelucrarea statistică a datelor înregistrate se obțin $\bar{x}_2 = 40,9$ şi $s_2^2 = 0,71$.

Material cu drept de autor

Se cere să se evalueze dacă cele două dispersii exprimă acelaşi câmp de dispersie general al calității caracteristicii considerate.

Rezolvare. Se determină raportul $F = s_1^2 / s_2^2$ =1,15. Pentru ν_1 = n_1- 1 = 30 – 1 = 29 şi ν_2 = n_2- 1 = 20 – 1 = 19 se găseşte $F_P = F_{0,95}$ = 2,07. Întrucât F = 1,15 < $F_{0,95}$= 2,07, nu există nici un temei să se respingă ipoteza că cele două dispersii de selecție s_1^2 şi s_2^2 exprimă aceeaşi dispersie generală σ^2.

6.3. Criterii de egalitate a unui şir de dispersii

Criteriul lui Cochran. În cazul comparării unui şir de dispersii de la analiza statistică a aceleiaşi caracteristici de calitate, de acelaşi volum de probe elementare, se foloseşte criteriul lui Cochran. Pentru aceasta se determină raportul:

$$G_{\max} = \frac{\left[s_i^2\right]_{\max}}{\sum_{i=1}^{m} s_i^2} \qquad (6.3)$$

în care $[s_i^2]_{\max}$ este dispersia cea mai mare din şirul de dispersii considerat; m – numărul de dispersii de selecție.

Dacă $G_{max} \leq G_q$ (anexa 6) se consideră că şirul de dispersii este omogen, adică exprimă aceeaşi dispersie generală σ^2.

Exemplul de calcul 6.2. După efectuarea a cinci extrageri, la momente diferite, a câte 20 probe elementare pentru o caracteristică de calitate şi determinarea dispersiilor de selecție se obțin valorile: s_1^2 = 1,54; s_2^2 = 2,08; s_3^2 = 1,86; s_4^2 = 1,97; s_5^2 = 1,58. Se cere să se determine dacă cele cinci dispersii exprimă aceeaşi dispersie generală σ^2.

Rezolvare. Se determină:

$$G_{\max} = \frac{\left[s_i^2\right]_{\max}}{\sum_{i=1}^{m} s_i^2} = \frac{2,08}{1,54 + 2,08 + 1,86 + 1,97 + 1,53} = 0,23$$

Din anexa 6, pentru **q = 0,05, m = 5** şi ν **= k – 1 = 20 – 1 = 19**, se obține $G_{0,05}$ **= 0,35**. Întrucât G_{max} **= 0,23 <** $G_{0,05}$ **= 0,35**, se poate spune că nu există nici un temei pentru a respinge ipoteza omogenității celor cinci dispersii, în sensul că ele exprimă aceeaşi dispersie generală σ^2.

Material cu drept de autor

6.4. Criterii ale egalității mediilor de selecție

Testul Student pentru două medii. Dacă s_1^2 şi respectiv s_2^2 exprimă aceeaşi dispersie generală σ^2, se trece la analiza condiției că $\overline{x}_1$ şi $\overline{x}_2$ exprimă aceeaşi valoare medie aritmetică generală $\mathbf{m_a}$.

Se va calcula dispersie s^2 cu relația:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \tag{6.4}$$

şi respectiv:

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{6.5}$$

Dacă $|t| \leq t_{\alpha,k}$ (anexa 3), atunci nu există temei a respinge ipoteza egalității celor două medii aritmetice, în sensul că exprimă aceeaşi medie generală $\mathbf{m_a}$.

Testul lui Dixon pentru verificarea omogenității a k valori medii aritmetice de acelaşi volum de selecție. Pentru fiecare probă (selecție) se calculează media. Mediile astfel obținute se ordonează crescător după cum urmează:

$$\mathbf{m_1 < m_2 < m_3 ; ... < M_3 < M_2 < M_1}$$

unde $\mathbf{m_1, m_2, m_3}$ sunt mediile cu valorile cele mai mici, iar $\mathbf{M_1, M_2, M_3}$ sunt mediile cu valorile cele mai mari.

Se calculează apoi rapoartele $\mathbf{r_{ij}}$ atât pentru cea mai mică medie, cât şi pentru cea mai mare, după modelul din tabelul 6.1.

Tabelul 6.1.

Tabel de calcul pentru testul Dixon

Raport	Testul pentru cea mai mică medie	Testul pentru cea mai mare medie
$\mathbf{r_{10}}$	$\frac{m_2 - m_1}{M_1 - m_1}$	$\frac{M_1 - M_2}{M_1 - m_1}$
$\mathbf{r_{11}}$	$\frac{m_2 - m_1}{M_2 - m_1}$	$\frac{M_1 - M_2}{M_1 - m_2}$
$\mathbf{r_{12}}$	$\frac{m_2 - m_1}{M_3 - m_1}$	$\frac{M_1 - M_2}{M_1 - m_3}$

Material cu drept de autor

r_{20}	$\frac{m_3 - m_1}{M_1 - m_1}$	$\frac{M_1 - M_3}{M_1 - m_1}$
r_{21}	$\frac{m_3 - m_1}{M_2 - m_1}$	$\frac{M_1 - M_3}{M_1 - m_2}$
r_{22}	$\frac{m_3 - m_1}{M_3 - m_1}$	$\frac{M_1 - M_3}{M_1 - m_3}$

Aceste rapoarte se compară cu cele teoretice ce rezultă din tabelul 6.2 determinate pentru un nivel de semnificație de 1%. În cazul când raportul calculat depăşeşte valoarea rezultată din tabele se trage concluzia că media de selecție (cea mai mare sau cea mai mică) diferă semnificativ, iar selecția respectivă se elimină ca necorespunzătoare. După eliminarea unei selecții, se repetă procedeul pentru eliminare (dacă este cazul) a selecției următoare.

Tabelul 6.2

Valorile r_{ij} pentru testul lui Dixon

Nr. selecții	r_{10}	r_{11}	r_{12}	r_{20}	r_{21}	r_{22}
4	0,926	0,995	-	0,996	-	-
5	0,821	0,937	0,996	0,950	0,998	-
6	0,740	0,839	0,951	0,865	0,970	0,998
7	0,680	0,782	0,875	0,814	0,919	0,970
8	0,634	0,725	0,797	0,746	0,868	0,922
9	0,598	0,677	0,739	0,700	0,816	0,873
10	0,568	0,639	0,694	0,664	0,760	0,826
11	0,542	0,606	0,658	0,627	0,713	0,781
12	0,522	0,580	0,629	0,612	0,675	0,740
13	0,503	0,558	0,612	0,590	0,649	0,705
14	0,488	0,539	0,580	0,571	0,627	0,674
15	0,475	0,522	0,560	0,554	0,607	0,647
16	0,463	0,508	0,544	0,539	0,589	0,624
17	0,452	0,495	0,529	0,526	0,573	0,605
18	0,442	0,484	0,516	0,514	0,559	0,589
19	0,433	0,473	0,504	0,503	0,547	0,575
20	0,425	0,464	0,493	0,494	0,536	0,562
21	0,18	0,455	0,483	0,485	0,526	0,551
22	0,411	0,446	0,474	0,477	0,517	0,541
23	0,404	0,439	0,465	0,469	0,509	0,532
24	0,399	0,432	0,457	0,462	0,501	0,524
25	0,393	0,426	0,450	0,456	0,493	0,516

Exemplul de calcul 6.3. Se consideră ca la analiza statistică prin metoda selecțiilor mici s-au înregistrat **25** probe selecții pentru care, în urma calculelor adecvate au rezultat valorile medii şi dispersiile din tabelul 6.3.

Material cu drept de autor

Rezolvare: După cum se observă, cea mai mică medie de sondaj este $\bar{x}_5 = 38,65$, iar cea mai mare medie este $\bar{x}_{21}$ = **44,67**. Prin urmare:

$$r_{10} = \frac{39,47 - 38,65}{44,67 - 38,65} \quad \text{şi} \quad r_{10} = \frac{44,67 - 44,35}{44,67 - 38,65} = 0,05$$

Din tabelul 6.3, pentru 22 de selecţii se obţine r_{10} = 0,411. Cum 0,12<0,411, respectiv 0,05<0,411, cu o probabilitate de 99% se poate afirma că atât cea mai mică medie de selecţie $\bar{x}_5 = 38,65$ cât şi cea mai mare medie de selecţie $\bar{x}_{21} = 44,67$ nu diferă semnificativ de celelalte. Rezultă:

$$m_a = \frac{1}{r}\sum_{i=1}^{22} \bar{x}_i = \frac{929,60}{22} = 42,45$$

Tabelul 6.3

Datele de calcul pentru exemplul 6.3

Nr. selecţiei	$\bar{x}_i$	s_i^2	Nr. selecţiei	$\bar{x}_i$	s_i^2
1	44,35	0,32	14	42,18	0,39
2	42,43	2,26	15	43,87	2,11
3	43,88	1,99	16	39,47	2,88
4	40,45	0,56	17	43,20	2,95
5	38,65	2,90	18	41,93	6,87
6	41,75	3,06	19	42,53	0,77
7	42,33	1,55	20	42,37	1,90
8	41,03	4,48	21	44,67	2,91
9	41,93	0,88	22	39,77	5,87
10	43,40	1,03	23	43,32	1,76
11	42,58	0,68	24	41,20	2,19
12	42,78	0,79	25	42,68	2,25
13	39,58	0,89			

Observaţii. S-au menţionat 22 medii aritmetice întrucât mediile nr. 18, 22 şi 8 au fost eliminate deoarece dispersiile acestora se abat semnificativ de la omogenitate (a se vedea analiza omogenităţii dispersiilor de selecţie).

La prima vedere, am putea fi tentaţi să rezolvăm problema prin compararea mediilor grupelor, prin metodele prezentate. Dar, există cel puţin trei argumente pentru care această opţiune nu este de dorit a fi urmată:

- În primul rând, volumul calculelor ar urma sa fie destul de mare şi ar creşte şi mai mult dacă numărul categoriilor variabilei independente ar fi din ce în ce mai mare.
- În al doilea rând, problema cercetării vizează relaţia dintre variabila dependentă (în exemplul de mai sus, performanţa la statistică) şi

Material cu drept de autor

variabila independentă, exprimată prin ansamblul tuturor categoriilor sale (grupele de studiu). Ar fi bine să putem utiliza un singur test şi nu mai multe, pentru a afla răspunsul la problema noastră.

• În fine, cel mai puternic argument, este acela că, prin efectuarea repetată a testului **t,** prin comparația mediilor două câte două, se acumulează o cantitate de eroare de tip **I** mai mare decât este permis pentru o decizie statistică (**0.05**), iar testul lui Dixon se aplică numai în cazul selecțiilor de acelaşi volum.

Pentru a se elimina aceste neajunsuri şi, mai ales pe ultimul dintre ele, se utilizează procedura statistică numită **analiza de varianță** (denumită pe scurt **ANOVA**, de la „***AN****alysis Of* ***VA****riance*", în engleză). În mod uzual, în esență, **ANOVA** nu este altceva decât o extensie a logicii testului ***t*** pentru situațiile în care se doreşte compararea a mai mult de două medii independente. Dar, dacă problema este similară, soluția este, aşa cum vom vedea, diferită.

Există mai multe tipuri de **ANOVA**, două fiind mai frecvent folosite:

ANOVA unifactorială:

• Presupune o variabilă dependentă măsurată pe o scală de interval/raport.

• Presupune o variabilă independentă de tip categorial (nominală sau ordinală) care ia trei sau mai multe valori. În contextul **ANOVA**, variabila independentă este definită ca „factor". Modelul de analiză de varianță cu o singura variabilă independentă se numeşte „**ANOVA unifactorială**" sau „**ANOVA simplă**".

ANOVA multifactorială

• Presupune o variabilă dependentă (la fel ca în cazul **ANOVA unifactorială**)

• Presupune două sau mai multe variabile independente, fiecare cu două sau mai multe valori măsurate pe o scală nominală sau ordinală.

6.5. Analiza dispersională (ANOVA) unifectorială

În cercetarea experimentală este frecventă comparația simultană între mediile a mai mult de două grupe, formate din subiecți supuşi la tratamente diferite sau cu date adunate în condiții diverse. Cu scopul de a evidenția toate posibilele diferențe semnificative între medii, nu este corect să se recurgă la testul **t** a lui Student pentru a repeta analiza de atâtea ori, câte comparații sunt posibile între perechi de grupe singulare.

Cu metoda **t** a lui Student, se utilizează numai o parte a datelor şi probabilitatea α aleasă pentru acceptarea ipotezei H_0, probabilitatea de a comite o eroare de speța I-a (excluderea ipotezei nule când în realitate este adevărată) e validă numai pentru orice singură comparație.

Dacă comparațiile sunt numeroase, probabilitatea complexă că cel puțin una dintre ele se dovedeşte semnificativă numai prin efectul cazului este mai mare. Dacă e adevărată ipoteza H_0 probabilitatea ca nici o

Material cu drept de autor

comparație să fie semnificativă este: **(1 - α)^n** - unde **n** este numărul de comparații efectuate.

De exemplu, dacă se efectuează 10 comparații între mediile din grupe extrase aleator din aceeaşi populație şi pentru fiecare din ele α = **0,05**, probabilitatea ca nici o comparație să rezulte semnificativ se micşorează la circa **0,60** (corespunde la **0,95^{10}**). Prin urmare, probabilitatea complexă ca cel puțin una să rezulte semnificativă numai prin efectul fluctuațiilor întâmplătoare devine **0,40**. Exprimat în termeni mai formali, efectuând **k** comparații cu testul **t** a lui Student, fiecare cu probabilitatea α, probabilitatea complexă α^* de a comite cel puțin o eroare de speța I-a (adică ca testul să refuze ipoteza nulă, când în realitate ea este adevărată) devine: $\alpha^* = 1-(1-\alpha)^k$.

În analiza variației, cu un aparent paradox ai termenilor, compararea este între două sau mai multe medii. Ea permite compararea simultană între ele, menținând constantă probabilitatea α complexă prefixată.

Ipoteza nulă **H_0** şi ipoteza alternativă **H_1** asumă o formulare mai generală, referitoare la compararea între două medii:

$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k,$

H_1 : nu sunt toate mediile egale (sau cel puțin una dintre μ_i diferă semnificativ de celelalte sau, cu alte cuvinte, cel puțin două μ_i sunt diferite între ele).

Metodologia dezvoltată pentru a verifica semnificația diferențelor între mediile aritmetice a diferitelor eşantioane de date experimentale, numită analiza variației, indicată prin **ANOVA**, utilizează distribuția **F**.

Este bazată pe raportul între variațe, numită test **F** de la Fisher (1890-1962), considerat cel mai eminent statistician contemporan şi tatăl statisticii moderne. Propunerea sa din 1925 permite să se descompună şi să se măsoare incidența diferitelor surse de variație asupra valorilor observate din două sau mai multe eşantioane. Este metodologia care stă la baza statisticii moderne; din ea au derivat progresiv analize mai complexe, cu care sunt luați în considerație acum mulți factori fie independenți fie corelați.

Metodologia actuală a **ANOVA** este totuşi datorată lui Snedecor (statistician american 1881-1974) care cu testul său scurt din 1934 a perfecționat metoda şi ia simplificat forma față de cea propusă original de Fisher. Lui Snedecor, împreună cu Cochran, i se datorează un alt test statistic care din 1934 până la ultima ediție din 1980 a fost timp de 50 ani un punct de referință fundamental pentru toți statisticienii. De aceea, distribuția **F** este menționată şi ca distribuție Fisher-Snedecor.

Marea revoluție introdusă de **ANOVA** referitoare la testul **t** constă în diferita aproximare în programarea experimentelor. Aproximarea testului **t** răspunde vechii axiome că **natura răspunde numai la întrebări simple**. Pentru organizarea unui experiment, materialul cu care se formează grupele de comparat trebuia să fie cel mai omogen posibil. De exemplu, pentru a compara efectul a două substanțe toxice asupra unui grup de dăunători

Material cu drept de autor

(insecte), dăunătorii trebuiau să fie de acelaşi sex, de aceeaşi vârstă, şi de aceeaşi dimensiune etc, dacă se considera că sexul, vârsta, greutatea, şi orice altă caracteristică cunoscută ar fi influențat asupra răspunsului experimentului.

Diferența între două eşantioane putea să rezulte mai uşor semnificativă, cu cât eroarea standard rezulta fără îndoială minoră; dar concluziunile erau limitate la grupul de dăunători cu caracteristici selecționate, fără posibilitatea de a le extinde la eşantioane cu caracteristici diferite. Pentru a face concluziile mai generale, nu rămânea decât să se repete experimentul variind pe rând câte un caracter. Era cerută o puternică creştere a cantității de materiale şi o prelungire a timpului necesar experimentului; la sfârşit, cu atâtea răspunsuri singulare, era foarte complex sau greu să se tragă concluzii generale.

Marea noutate introdusă de **ANOVA** este descoperirea avantajelor oferite analizelor de folosirea unui material foarte diversificat. Cunoscând cauzele şi diferiții factori, este posibil să se atribuie fiecăruia dintre aceştia efectul său şi să se reducă variabilitatea erorii. Diferențele între mediile eşantioanelor devin mult mai uşor semnificative şi concluziile pot fi imediat extinse situațiilor variate.

De la introducerea **ANOVA**, în programarea şi realizarea unui experiment este avantajos să folosim un material neomogen pentru toate caracterele, ci numai pentru unele.

În **ANOVA**, sursa sau cauza variațiilor datelor se numeşte factor experimental sau tratament. El poate să fie:

- cu mai multe nivele cantitative, ca dozele crescătoare a aceleiaşi substanțe;
- cu diverse modalități calitative, ca administrarea substanțelor diferite.

Orice unitate de observație a grupului experimental se numeşte replică; pentru a permite calcularea mediei şi a variației, orice grup (eşantion) trebuie să fie format din două replici.

6.5.1. Analiza variației cu un criteriu de clasificare şi eşantionare randomizată

Modelul cel mai simplu de analiză a variației, care poate fi privită ca o extensie a testului **t** a lui Student, la mai multe probe independente, e numit cu un criteriu de clasificare (unifactorială): orice dată este clasificată numai pe baza eşantionului căruia aparține. Se numeşte şi model complet randomizat întrucât, mai ales în analiza de laborator prevede o eşantionare în care **n** indivizi omogeni aparțin întâmplător la diferite nivele ale factorului.

Când se dispune de un grup de subiecți pentru a fi supuşi diverselor tratamente ca să se compare efectele, atribuirea fiecărui exemplar un tratament specific, trebuie să aibă loc prin extracție întâmplătoare din întregul grup.

Material cu drept de autor

Metodologia de prezentare a observațiilor, codificată, prevede că datele experimentale culese să fie raportate în mod ordonat după tabelul 6.4. Pentru analiza statistică în acest model nu este cerut ca diferitele eşantioane să aibă acelaşi număr de observații ($\mathbf{n_i}$) sau de replici.

Observația singulară $\mathbf{X_{ij}}$ este raportată cu doi indici relativi (1 relativ la grup sau tratament şi 2 la poziția ocupată în grup).

Media fiecărui grup sau a unui singur tratament $\overline{X}_{\bullet i}$ este prezentată supraliniată cu o linie şi cu indicele de grup.

Media generală $\overline{\overline{X}}$ a tuturor datelor este indicată dublu barat şi fără indici.

Plecând de la aceste 3 cantități se estimează abaterile şi variațele utile analizei.

ANOVA e bazată pe efectele aditive a diferiților factori considerați. În modelul cel mai simplu, care consideră numai un factor la două sau mai multe nivele, fiecare singură observație $\mathbf{X_{ij}}$ poate să fie scrisă prin expresia:

$$\mathbf{X_{ij}} = \mu + \alpha_i + \varepsilon_{ij} \qquad (6.6)$$

În care: μ - media generală, defineşte dimensiunea experimentului; α_i este factorul de tratament; ε_{ij} factor cauzal, numit rezidiu sau eroare experimentală (este important să amintim că eroare nu este sinonim cu greşeală, ci indică efectul unuia sau mai multor factori necunoscuți, oricum neevaluați sau necontrolați în experiment).

Tabelul 6.4

Înregistrarea datelor experimentale pentru analiza dispersională unifactorială

Unități experimentale sau replici	Modalități sau nivele de tratamente				
	T_1	T_2	T_3	...	T_p
1	X_{11}	X_{12}	X_{13}	...	X_{1p}
2	X_{21}	X_{22}	X_{23}	...	X_{2p}
3	X_{31}	X_{32}	X_{33}	...	X_{3p}
...	...	...	...	...	...
n_i	$X_{n_1 1}$	$X_{n_2 2}$	$X_{n_2 3}$	...	$X_{n_p p}$
Media tratamentelor	$\overline{X}_{\bullet 1}$	$\overline{X}_{\bullet 2}$	$\overline{X}_{\bullet 3}$	...	$\overline{X}_{\bullet p}$
Media generală	$\overline{\overline{X}}$				

La un exemplu, cu trei grupe de subiecți (**A, B, C**) cărora le-a fost măsurată o caracteristică cantitativă a unei substanțe, în mg, s-au obținut rezultatele prezentate în tabelul 6.5.

Însă, datele trebuie să fie citite ca şi cum ar fi fost scrise în modul în care sunt prezentate în tabelul 6.6.

Reprezentarea grafică a valorilor observate ilustrează cu claritate, încă mai mare, conceptul.

Material cu drept de autor

În figura 6.1: linia orizontală centrală continuă reprezintă media generală; cele trei linii orizontale întrerupte reprezintă mediile celor trei eşantioane; punctele marcate reprezintă observaţiile singulare.

Tabelul 6.5.

Replici	A	B	C
1	**2,4**	**3,2**	**2,1**
2	**2,7**	**2,9**	**2,7**
3	**2,7**	**3,5**	**2,7**
4	**2,6**	---	---
Media	**2,6**	**3,2**	**2,5**

Tabelul 6.6

Replici	A	B	C
1	**2,6-0,2**	**3,2+0,0**	**2,5-0,4**
2	**2,6+0,1**	**3,2-0,3**	**2,5+0,2**
3	**2,6+0,1**	**3,2+0,3**	**2,5+0,2**
4	**2,6+0,0**	---	---
Media	**2,6**	**3,2**	**2,5**

Punctele indicate apar mai puţin numeroase decât datele, pentru că unele valori sunt egale şi deci punctele sunt suprapuse. Din cauza programului, grupele **A, B, C** în grafic sunt indicate cu cifrele **1, 2, 3**.

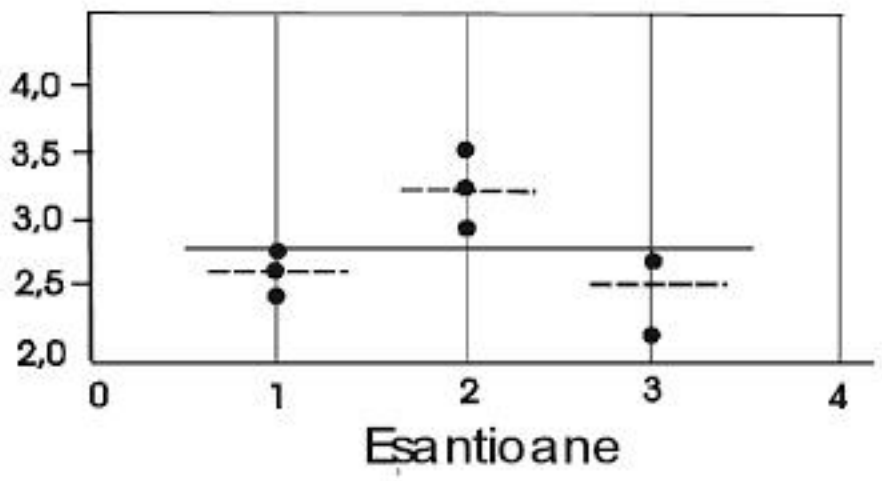


Fig.6.1. Reprezentarea grafică a datelor din tabelul 6.6

Într-un astfel de model, efectul α al tratamentului este la rândul său măsurat ca fiind: $\alpha = \mu_i - \mu$ unde μ_i este media eşantionului şi μ media generală.

Trecând de la enunţul teoretic la datele experimentale, se poate scrie că orice dată singulară $\mathbf{X}_{ij}$ a unui tratament specific poate fi:

$$X_{ij} = \overline{\overline{X}} + (\overline{X}_i - \overline{\overline{X}}) + \varepsilon_{ij} \qquad (6.7)$$

Conform expresiei (6.7) valorile singulare sunt determinate de media generală $\overline{\overline{X}}$, de efectul tratamentului pe fiecare eşantion $(\overline{X}_i - \overline{\overline{X}})$, de alţi factori necunoscuţi, simbolizaţi prin ε_{ij}.

Înainte de aplicarea acestui test parametric, trebuie să verificăm dacă există condiţii pentru el.

Presupunerile de **validitate a testului F** depind de erorile ε_{ij} care:

- trebuie să fie **independente** între ele;
- trebuie să fie **distribuite normal**;
- varianţele diferitelor grupări (eşantioane) trebuie **să fie omogene**.

Independenţa erorilor comportă ca variaţia întâmplătoare a oricărei observaţii să nu fie influenţată de variaţia unei alteia: eroare unei replici, abaterea sa faţă de media grupului de apartenenţă, nu trebuie să fie

Material cu drept de autor

influențată nici de semnul (când se pot avea valori pozitive şi negative) nici de mărimea valorii sale.

În acest scop randomizarea trebuie să fie bazată pe elemente obiective (efectul Random) şi nu lăsată arbitrar la intuiția experimentatorului; orice dată trebuie să aibă aceeaşi posibilitate de a fi influențată de factorii cunoscuți (efectul de tratament) şi de cei necunoscuți (efectul ambiant statistic).

Erorile trebuie să fie distribuite normal în jurul mediei. Înainte de aplicarea testului trebuie să fie făcut controlul asimetriei şi boltirii distribuției, pentru a verifica dacă nu se depărtează excesiv de normală. Când depărtarea este semnificativă, adesea este posibil să se reconstruiască condițiile de validitate prin transformarea datelor (care vor fi prezentate succesiv).

Omogenitatea variantei, prin care diferitele grupe din care se compară respectivele medii trebuie să aibă toate aceeaşi varianță adevărată (σ^2), este indispensabilă pentru a nu determina pierderi de informație prin efectul tratamentelor. Şi în acest caz, poate să fie necesară transformarea datelor.

După analiza datelor pentru verificarea condițiilor de validitate, metodologia analizei variantei prevede calculul următoarelor cantități:

- **abaterea totală**, cu ale sale grade de libertate,
- **deviația între grupele de date statistice**, cu gradele sale de libertate şi varianța relativă respectivă;
- **deviația intra grupe** cu gradele lor de libertate şi varianța relativă respectivă.

La sfârşitul unei verificări a rezultatelor şi a elaborărilor lor succesive, este util să amintim că suma abaterii între grupele de date statistice şi aceea intra grupe este egală cu abaterea totală; o proprietate identică aditivă au şi respectivele grade de libertate.

Abateri, gradele de libertate şi variantele unei analize a variantei sunt obişnuit prezentate ca în tabela următoare:

Rezultatele analizei dispersionale se prezintă în tabelul 6.7.

Tabelul 6.7

Componenta dispersională (abaterea)	Numărul gradelor de libertate	Dispersia
Între grupele de date statistice	$\upsilon_1 = \mathbf{n - 1}$ (**n** – nr. datelor din eşantion)	s_1^2
În interiorul grupelor (reziduală)	$\upsilon_2 = \mathbf{p - 1}$ (**p** – nr. eşantioane)	s_2^2
Generală	$\upsilon_3 = \mathbf{n - p}$	s^2

Raportul dispersiilor:

Abaterea totală sau SQ_{total} (suma pătratelor abaterilor, engleză - **SS**) este calculată din:

Material cu drept de autor

$$SQ_{total} = \sum_{j=1}^{p}\sum_{i=1}^{n_j}(X_{ij} - \overline{\overline{X}})^2 = \sum_{j=1}^{p}\sum_{i=1}^{n_j}X_{ij}^2 - \frac{\left(\sum_{j=1}^{p}\sum_{i=1}^{n_j}X_{ij}\right)^2}{n} \qquad (6.8)$$

Prima relație este numită **formula euristică**, deoarece definește semnificația abaterii totale: **suma pătratelor abaterilor de orice valoare de la media generală**.

A doua relație este **formula simplificată, matematic echivalentă cu prima**, care face mai simple și mai rapide calculele necesare. Cu aceasta, **abaterea totală este obținută ca diferența între suma pătratelor tuturor datelor și pătratul sumei tuturor datelor împărțit la numărul de date**.

A doua formulă are avantajul că cere mai puține operații și că nu utilizează media care adesea este o valoare aproximată; în aceste condiții conține un calcul mai precis decât formula euristică.

Abaterea între grupe (**$SQ_{între}$**) este prin definiție (**formulă euristică**) suma pătratelor diferențelor între orice medie a grupului și media generală, înmulțit cu numărul de date al grupului respectiv.

$$SQ_{\text{int}re} = \sum_{j=1}^{p} n_i(\overline{X}_j - \overline{\overline{X}})^2 = \sum_{j=1}^{p}\frac{\left(\sum_{i=1}^{n_j}X_i\right)^2}{n_i} - \frac{\left(\sum_{j=1}^{p}\sum_{i=1}^{n_j}X_{ij}\right)^2}{n} \qquad (6.9)$$

Formula prescurtată utilizează suma grupelor și suma totală, determinând cu precizie mai mare a rezultatelor.

Abaterea în interiorul grupelor (**SQ_{intro}**), numită și eroare.

$$SQ_{\text{int}ro} = \sum_{j=1}^{p}\sum_{i=1}^{n_j}(X_{ij} - \overline{X}_j)^2 = SQ_{total} - SQ_{\text{int}re} \qquad (6.10)$$

Este suma abaterilor la pătrat a oricărei valori de la media grupului său.

Prin proprietatea aditivă a abaterilor, poate fi obținută scăzând din abaterea totală, abaterea între grupe.

Gradele de libertate sunt determinate din numărul de sume cerute în calculul abaterilor relative în formula euristică.

- pentru abaterea totală, unde însumarea este extinsă la toate cele **n** date, **gdl** sunt **n-1**.
- pentru abaterea între tratamente, unde însumarea este extinsă la **p** grupe, **gdl** sunt **p-1**.
- pentru abaterea în intra sau eroarea, însumarea este extinsă la toate datele din interiorul fiecărui grup. Pentru a calcula gradele de libertate trebuie deci să scădem **1** din datele oricărui grup și deci este determinată de **n-p**. Prin proprietatea aditivă a **gdl**, se poate scrie și că [**(n-1)-(p-1)**], care dă **n-p**.

Material cu drept de autor

Împărțind abaterea între grupe şi aceea în interiorul grupelor la respectivele **gdl**, se obțin abaterea între şi abaterea intra (varianța totală e lipsită de interes la sfârşitul acestui test).

Varianța între grupe măsoară diferențele existente între un grup şi altul, chiar dacă calculul este făcut în raport cu media generală. Varianța în interiorul grupelor măsoară variabilitatea existentă în jurul mediei aritmetice a oricărui grup.

Dacă e adevărată ipoteza nulă, datele diferitelor grupe sunt extrase întâmplător din aceeaşi populație. Varianța între mediile grupelor şi varianța în interiorul oricărui grup depind de variabilitatea existentă între date: varianța între (s_F^2) şi variația intra s_e^2 sunt două estimări independente a aceleiaşi variante adevărate σ^2 şi deci ar trebui să aibă statistic aceeaşi valoare.

Ca indice al egalității între cele două variante, este folosit testul **F** a lui Fisher bazat pe raportul: **varianța între/varianța intro**, adică:

$$F_{(p-1,n-p)} = \frac{s_F^2}{s_e^2} \qquad (6.11)$$

Dacă este adevărată ipoteza nulă $\mathbf{H_0}$:

$H_0 : \mu_1 = \mu_2 = ... = \mu_k$, raportul ar trebui să rezulte **egal cu 1**.

Dacă este adevărată ipoteza alternativă $\mathbf{H_1}$:

H_1 : nu toate mediile sunt egale (sau cel puțin una dintre μ_i diferă semnificativ de celelalte sau, cu alte cuvinte, cel puțin două μ_i sunt diferite între ele), raportul ar trebui să rezulte **superior lui 1**.

Testul şi tabela respectivă sunt unilaterale, tocmai pentru că valoarea trebuie să fie mai mare decât 1.

Cu un număr infinit de grupe şi replici este suficient un raport superior lui **1** pentru a exclude ipoteza nulă (cum arată tabela valorilor critice ale lui **F**); cu un număr redus de date, raportul poate să fie superior lui 1 prin efectul variațiilor accidentale.

Valorile critice pentru respectivele grade de libertate sunt date de distribuția **F**.

- dacă valoarea lui **F** calculat este superioară celei tabelate, la o probabilitate α fixată, se exclude ipoteza nulă şi se acceptă ipoteza alternativă: cel puțin o medie este diferită de celelalte.
- Dacă valoarea **F** calculată este inferioară celei raportate în tabelă, se acceptă ipoteza nulă sau cel puțin nu poate să fie exclusă pentru că mediile sunt toate egale.

Exemplu. Pentru un control al calității aerului, cu prelevări din trei zone diferite ale unui oraş (numite **A, B, C**) a fost măsurată şi cantitatea de **Fe** (în micrograme/N_mc, la 0^0C şi 1013 mbar) între metalele grele în suspensie, tabelul 6.8.

Material cu drept de autor

Întrebare. Există o diferență semnificativă între cele trei zone în ceea ce priveşte cantitatea de fier în suspensie ?

Răspuns. Ipoteza nulă H_0 este că între mediile celor trei probe nu există diferențe semnificative.

Ipoteza nulă: H_0: $\mu_A = \mu_B = \mu_C$.

Ipoteza alternativă H_1: nu toate μ_i sunt toate egale.

Prin testul **F** este posibil să estimăm probabilitatea de a găsi între abateri medii egale sau superioare celor experimentale observate, în ipoteza că H_0 este adevărată.

Ca prim pas, dintre cele trei serii de date trebuie să calculăm:

- totalul fiecărei coloane: ΣX_j;
- numărul de observații: n_j;
- media fiecărei coloane: $\bar{X}_j$,

Tabelul 6.8

Factor experimental		
A	B	C
2,71	1,75	2,22
2,06	2,19	2,38
2,84	2,09	2,56
2,97	2,75	2,60
2,55	---	2,72
2,78	---	---

Succesiv, din acestea este necesar să estimăm:

- suma totală: ΣX,
- numărul total de observații: **N**,
- media totală sau generală: $\bar{\bar{X}}$.

Cum se arată în tabela următoare, 6.9:

Tabelul 6.9

	A	B	C		
ΣX_j	15,91	8,78	12,48	ΣX	37,17
n_j	6	4	5	N	15
$\bar{X}_j$	2,652	2,195	2,496	$\bar{\bar{X}}$	2,478

Plecând de la calculele din tabelul 6.9, se calculează abaterile şi numărul gradelor de libertate respective.

Abaterea totală poate să fie calculată din suma pătratelor abaterilor fiecăreia dintre cele **15** observații față de media generală, în acord cu formula euristică:

$$SQ_{total} = \sum_{j=1}^{p} \sum_{i=1}^{n_j} (X_{ij} - \bar{\bar{X}})^2 \qquad (6.12)$$

Tabelul 6.10

A	B	C
$(2,71 - 2,478)^2$	$(1,75 - 2,478)^2$	$(2,22 - 2,478)^2$
$(2,06 - 2,478)^2$	$(2,19 - 2,478)^2$	$(2,38 - 2,478)^2$

Material cu drept de autor

$(2,84 - 2,478)^2$	$(2,09 - 2,478)^2$	$(2,56 - 2,478)^2$
$(2,97 - 2,478)^2$	$(2,75 - 2,478)^2$	$(2,60 - 2,478)^2$
$(2,55 - 2,478)^2$	---	$(2,72 - 2,478)^2$
$(2,78 - 2,478)^2$	---	---

Dezvoltând calculele şi însumând rezultatele se obțin datele din tabelul 6.11

Tabelul 6.11

A	B	C
0,053824	0,529984	0,066564
0,174724	0,082944	0,009604
0,131044	0,150544	0,006724
0,242064	0,073984	0,014884
0,005184	---	0,058564
0,091204	---	---
Σ: 0,698040	Σ: 0,837456	Σ: 0,156340

Abaterea totală este:

$$SQ_{total} = 0,698040 + 0,837456 + 0,156340 = 1,691836$$

cu 14 grade de libertate.

Această metodă de calcul a abaterii totale este lungă şi determină estimări neprecise când media generală este aproximată. Pentru calculul manual totdeauna este convenabil să se utilizeze formula prescurtată:

$$SQ_{total} = \sum_{j=1}^{p}\sum_{i=1}^{n_j} X_{ij}^2 - \frac{\left(\sum_{j=1}^{p}\sum_{i=1}^{n_j} X_{ij}\right)^2}{n} \qquad (6.13)$$

Care aplicată la datele exemplului:

Tabelul 6.12

A	B	C
$2,71^2 = 7,3441$	$1,75^2 = 3,0625$	$2,22^2 = 4,9284$
$2,06^2 = 4,2436$	$2,19^2 = 4,7961$	$2,38^2 = 5,6644$
$2,84^2 = 8,0656$	$2,09^2 = 4,3681$	$2,56^2 = 6,5536$
$2,97^2 = 8,8209$	$2,75^2 = 7,5625$	$2,60^2 = 6,7600$
$2,55^2 = 6,5025$	---	$2,72^2 = 7,3984$
$2,78^2 = 7,7284$	---	---
ΣX^2: 42,7051	ΣX^2: 19,7892	ΣX^2: 31,3048
$\Sigma X^2_{total} = 42,7051 + 19,7892 + 31,3048 = 93,7991$		

Material cu drept de autor

Din cele două diferite sume se estimează abaterea totală:

$$SQ_{total} = 93,7991 - \frac{(37.17)^2}{15} = 1,69184$$

Corespondența dintre cele două estimări este o demonstrație elementară şi intuitivă a echivalenței matematice ale celor două formule (diferența se datorează rotunjirilor).

Abaterea între diferite eşantioane măsoară variabilitatea existentă între media aritmetică a fiecărui eşantion şi media aritmetică generală, ponderată prin numărul de observații prezente în fiecare eşantion. Dacă n-ar exista variabilitate întâmplătoare, şi valoarea observațiilor singulare ar fi determinată numai de factorul specific care le regrupează, replicile fiecărui eşantion ar trebui să aibă toate aceeaşi valoare şi să fie egale cu media eşantionului, cum evidențiază formula euristică:

$$SQ_{\text{int re}} = \sum_{j=1}^{p} n_i (\overline{X}_j - \overline{\overline{X}})^2 \qquad (6.14)$$

Abaterea între eşantioane este suma abaterilor fiecărei medii din eşantionul respectiv de media generală, ponderată prin numărul de replici. Prin aceasta, cu formula euristică calculul devine:

$$\mathbf{SQ_{intre} = 6.(2,652-2,478)^2+4.(2,195+2,478)^2+5.(2,496-2,478)^2=}$$
$$\mathbf{= 6.0,030276 + 4.0,080089 + 5.0,000324 =}$$
$$\mathbf{= 0,181656 + 0,320356 + 0,00162 = 0,503632}$$

Şi în acest caz, formula prescurtată:

$$SQ_{\text{int re}} = \sum_{j=1}^{p} \frac{\left(\sum_{i=1}^{n_j} X_i\right)^2}{n_i} - \frac{\left(\sum_{j=1}^{p}\sum_{i=1}^{n_j} X_{ij}\right)^2}{n} \qquad (6.15)$$

este mai rapidă şi precisă, ne cerând aproximarea mediilor:

$$SQ_{\text{int re}} = \frac{(15,91)^2}{6} + \frac{(8,78)^2}{4} + \frac{(12,48)^2}{5} - \frac{(37,17)^2}{15} =$$
$$= 92,610196 - 92,10726 = 0,502936$$

Şi în acest caz diferențele sunt minime şi datorate utilizării unui număr diferit de zecimale în diferitele aproximații (obişnuit sunt suficiente calcule cu 2 sau 3 cifre zecimale; numărul cel mai mare utilizat aici este motivat din necesitatea de a confrunta rezultatele celor două metode).

Abaterea în interiorul eşantioanelor (**SQ_{intro}**), numită şi eroare,

$$SQ_{\text{int ro}} = \sum_{j=1}^{p}\sum_{i=1}^{n_j} (X_{ij} - \overline{X}_j)^2 \qquad (6.16)$$

măsoară variația între valoarea fiecărei replici şi media aritmetică a eşantionului respectiv.

Material cu drept de autor

Însumând aceste diferențe ridicate la pătrat pentru orice grup,

Tabelul 6.13

A	B	C
$(2,71 - 2,652)^2$	$(1,75 - 2,195)^2$	$(2,22 - 2,496)^2$
$(2,06 - 2,652)^2$	$(2,19 - 2,195)^2$	$(2,38 - 2,496)^2$
$(2,84 - 2,652)^2$	$(2,09 - 2,195)^2$	$(2,56 - 2,496)^2$
$(2,97 - 2,652)^2$	$(2,75 - 2,195)^2$	$(2,60 - 2,496)^2$
$(2,55 - 2,652)^2$	---	$(2,72 - 2,496)^2$
$(2,78 - 2,652)^2$	---	---

Dezvoltând calculele şi însumând rezultatele se obțin datele din tabelul 6.11

Tabelul 6.11

A	B	C
0,003364	0,198025	0,076176
0,350464	0,000025	0,013456
0,035344	0,011025	0,004096
0,101124	0,308025	0,010816
0,010404	---	0,050176
0,015376	---	---
Σ: 0,516076	Σ: 0,517100	Σ: 0,154720

După însumare rezultă:

$\mathbf{SQ_{intro} = 0,516076 + 0,517100 + 0,154720 = 1,187896}$

cu 12 grade de libertate.

Abaterea **intro** poate să fie obținută mult mai rapid prin scăderea abaterii **între** din abaterea **totală** calculată precedent.

$\mathbf{SQ_{intro} = SQ_{total} - SQ_{intre} = 1,69184 - 0,502936 = 1,88904}$

În acelaşi mod, prin proprietatea aditivă, se poate calcula numărul de grade de libertate:

$\mathbf{gdl_{intro} = gdl_{total} - gdl_{intre} = 14 - 2 = 12}$

Pentru o prezentare clară şi sintetică valorile calculate sunt rezumate într-o tabelă care dă cele trei abateri, respectivele grade de libertate (**gdl**) şi variantele utile testului.

Abaterea		gdl	Varianța
Totală	1,69184	14	---
Între eşantioane	0,502936	2	0,251468
În interiorul eşantioanelor	1,188904	12	0,0990753

Împărțind abaterea între şi abaterea în, prin respectivele **gdl**, se obțin varianța între şi varianța în interiorul eşantioanelor.

Împărțind varianța între prin varianța în, se calculează raportul F care trebuie redat cu respectivele **gdl** ($\mathbf{F_{2;\,12}}$)

Material cu drept de autor

$$F_{2;12} = \frac{0,251468}{0,0990753} = 2,538$$

Valoarea critică a lui **F** cu **2** grade de libertate pentru numărător şi **12** grade de libertate pentru numitor care este raportată în tabelă pentru probabilitatea α = **0,05** este **3,89**. Valoarea calculată (**2,538**) este inferioară celeia din tabel: probabilitatea ca ipoteza nulă să fie adevărată este superioară lui **5%**. În consecință se acceptă ipoteza nulă: cele trei probe sunt extrase din aceeaşi populație; nu există o diferență semnificativă între cele trei medii ale probelor.

6.5.2. Comparație între testul F de analiza variației cu două grupe de date statistice şi testul t Student pentru două probe independente

Analiza variației poate să fie aplicată şi pentru numai **2** grupe de date statistice; pentru acest caz a fost deja prezentată metodologia testului **t** Student. În realitate, testul **t** şi testul **F** sunt **2** moduri numai aparent diferite pentru a face aceeaşi analiză: testul **t** poate fi privit ca un caz special al analizei variației, aplicată numai la **2** eşantioane; mai mult, **analiza variației este extinderea la mai multe grupe şi la mai mulți factori a testului t Student**.

În cazul unui singur factor cu **2** eşantioane, între **t** şi **F** există o relație matematică precisă:

$$F_{1;\nu} = t_{\nu}^{2} \qquad (6.17)$$

care, evident poate să fie scrisă şi sub forma:

$$t_{\nu} = \sqrt{F_{1;\nu}} \qquad (6.18)$$

unde υ este numărul gradelor de libertate.

Valoarea lui F cu 1 grad de libertate la numărător şi ν la numitor este egală cu pătratul lui t cu υ grade de libertate.

Cele două distribuții a valorilor critice pentru aceeaşi probabilitate α sunt echivalente, aşa cum este posibil să se evidențieze din simpla comparare între tabelele valorilor critice.

Exemplu de calcul 6.4. Două eşantioane de câte 10 pui nou născuți, extrase întâmplător din aceeaşi populație, au fost crescute în două incinte separate cu două rețete noi de hrană, pentru a se verifica eficiența acestora asupra creşterii.

După o lună sunt cântăriți indivizii supraviețuitori: 7 din eşantionul A şi 8 din eşantionul B, cu valorile înregistrate în tabelul 6.12.

Tabelul 6.12

A	2,7	2,8	2,9	2,5	2,6	2,7	2,8	---
B	2,2	2,1	2,2	2,3	2,1	2,2	2,3	2,6

Material cu drept de autor

Reprezentarea grafică evidențiază caracteristicile celor 2 serii de observații (unele valori sunt identice şi deci punctele apar mai puțin numeroase decât datele pentru că sunt suprapuse). Din cauza programului, grupele A şi B în grafic sunt indicate respectiv cu 1 şi 2).

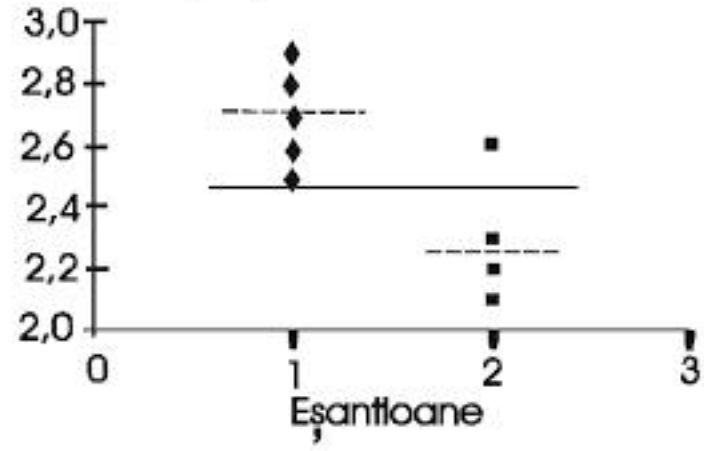


Fig.6.2. Reprezentarea grafică a datelor din tabelul 6.12

Rezolvare. Ipoteza nulă este:

H_0: $\mu_A = \mu_B$;

Ipoteza alternativă bilaterală este:

H_1: $\sigma^2_A \neq \sigma^2_B$.

Înainte de a proceda fie la testul **t** fie la testul **F**, trebuie să se verifice dacă cele 2 variante sunt omogene. Deci, preliminar se fac comparații între cele două medii şi comparația între cele două variante.

În acest scop se calculează cele două abateri şi gradele de libertate, pentru a estima variantele respective, tabelul 6.13:

Tabelul 6.13

	A	B
Abaterea	0,10857	0,18000
Grade de libertate	6	7
Varianța – s^2	0,018095	0,02571

Se efectuează raportul **F** între:

- varianța majoră la numărător;
- varianța minoră la numitor.

$$F_{7;6} = \frac{0,02571}{0,018095} = 1,42$$

În tabela valorilor critice pentru funcția Fisher-Snedecor, cu 7 grade delibertate pentru varianța majoră de la numărător şi 6 grade de libertate pentru varianța minoră de la numitor, valoarea critică la probabilitatea α = **0,05** este egală cu **4,21**. Valoarea calculată (**1,42**) este inferioară: în consecință se acceptă ipoteza nulă că cele două variante sunt omogene.

În continuare este corect să se procedeze la compararea celor două medii.

Pentru testul **t** Student, pentru două probe independente.

Se calculează cele două medii:

Media eşantionului **A = 2,714**,

Media eşantionului **B = 2,250**.

Şi ca urmare, varianța mediată:

$$s_p^2 = \frac{0,10825 + 0,18000}{6+7} = 0,022173$$

Din aceasta se estimează valoarea lui **t** cu 13 grade de libertate.

Material cu drept de autor

$$t_{13} = \frac{2,714 - 2,250}{\sqrt{0,022173 \times \left(\frac{1}{7} + \frac{1}{8}\right)}} = 6,02$$

Prin analiza variantei la un criteriu de clasificare, trebuie să se calculeze abaterea totală, abaterea **între** grupe şi **în** interiorul grupelor de date experimentale, cu respectivele grade de libertate.

Este posibilă o verificare a calculelor efectuate, prin proprietatea aditivă a abaterilor:

devianța totală = devianța între grupe + devianța în interiorul grupelor

Tabelul 6.14

	Abaterea	gdl	Varianța
Totală	1,093333	14	---
Între grupe	0,804762	1	0,804761
În interiorul grupelor	0,288571	13	0,022198

Se calculează varianța între şi varianța în şi din ele se estimează **F** cu 1 şi 13 **gdl**:

$$F_{(1;13)} = \frac{0,804761}{0,022198} = 36,25$$

rezultă egală cu 36,25.

E simplu de verificat că cele două răspunsuri coincid.

$$t^2_{(13)} = F_{(1,13)};\ \ (6,02)^2 = 36,25$$

Mai puțin aproximările determinate de nr. de zecimale.

În tabelele valorilor critice a testului **t** Student şi a testului **F** (Fisher) se controlează probabilitatea, care pentru amândouă rezultă egale şi net inferioare lui 0,001.

Cu amândouă testele se exclude ipoteza nulă cu aceeaşi probabilitate.

6.5.3. Teste pentru omogenitatea variantei între mai multe eşantioane: testele Hartley, Cochran, Bartlett

Compararea între medii cu analiza variantei cere ca diferitele grupe (eşantioane) să aibă variante egale. Îndepărtându-se sensibil de această condiție de validitate se influențează grav varianta erorii, adică semnificația testului. S-ar utiliza o variantă a erorii medii s^2, ca estimare a variantei adevărate σ^2, care ar rezulta prea mare pentru unele eşantioane şi prea mică pentru altele. În afară de verificarea condițiilor de validitate, pentru compararea între medii, adesea există şi un interes explicit pentru o comparare între variante. De exemplu,

Material cu drept de autor

- grupe de animale sau plante genetic identice ar trebui să aibă varianțe semnificative mai mici decât grupele genetic eterogene;

- grupuri de animale sau vegetale crescute în condiții de mediu foarte diferite ar trebui să aibă o varianță mai mare decât grupurile crescute în condiții similare;

- în analizele de laborator, un instrument de măsură mai precis sau un reactiv de calitate superioară ar trebui să furnizeze varianțe mai mici față de instrumentele şi reactivii de joasă calitate, în experimente repetate în aceleaşi condiții.

Ipoteza omogenității varianțelor în cazul mai multor grupe cere verificarea ipotezei nule:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \ldots = \sigma_p^2$$

Ipoteza alternativă:

$$H_1 : nu\ toate\ \text{var}ianțanțele\ \sigma_i^2\ sunt\ egale.$$

Metodele propuse sunt numeroase; între cele mai răspândite, utilizate şi în programele informatice standard pentru calculatoare sunt de amintit:

A. Testul $\mathbf{F_{max}}$ al lui Hartley;
B. Testul pentru varianța maximă şi varianța minimă a lui Cochran;
C. Testul lui Bartlett;

A. Procedeul $\mathbf{F_{max}}$ a lui Hartley este cel mai simplu şi rapid, fiind o generalizare a testului pentru două probe (grupe, eşantioane). Dificultățile în utilizarea lui derivă numai din redusa referire a testului la o amplă difuziune pe care o raportează tabela valorilor critice, tabelul 6.15.

Tabelul 6.15

Valori critice pentru testul Hartley de omogenitate a varianței dintre k grupe (eşantioane, probe), $\alpha = 0{,}05$.

Df_2	2	3	4	5	6	7	8	9	10	11	12
2	39,0	87,5	142	202	266	333	403	475	550	626	704
3	15,4	27,8	39,2	60,7	62,0	72,9	83,5	93,9	104	114	124
4	9,60	15,5	20,6	26,2	29,5	33,6	37,5	41,1	44,6	48,0	51,4
5	7,15	10,3	13,7	16,3	18,7	20,8	22,9	24,7	26,5	28,2	29,9
6	5,82	8,38	10,4	12,1	13,7	15,0	16,3	17,5	18,6	19,7	20,7
7	4,99	6,94	8,44	9,70	10,8	11,8	12,7	13,5	14,3	15,1	15,8
8	4,43	6,00	7,18	8,12	9,03	9,78	10,5	11,1	11,7	12,2	12,7
9	4,03	5,34	6,31	7,11	7,80	8,41	8,95	9,45	9,91	10,3	10,7
10	3,72	4,85	5,67	6,34	6,92	7,42	7,87	8,28	8,66	9,01	9,34
12	3,28	4,16	4,79	5,30	5,72	6,09	6,42	6,72	7,00	7,25	7,48
15	2,86	3,54	4,01	4,37	4,68	4,95	5,19	5,40	5,59	5,77	5,93

Material cu drept de autor

20	2,46	2,95	3,29	3,54	3,76	3,94	4,10	4,24	4,37	4,49	4,59
30	2,07	2,40	2,61	2,78	2,91	3,02	3,12	3,21	3,29	3,36	3,39
60	1,67	1,85	1,96	2,04	2,11	2,17	2,22	2,26	2,30	2,33	2,36
∞	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00

Să nu se confunde cu tabela Fisher-Snedecor, prezentă în testele respective (specifice). Ele coincid numai în cazul a două probe independente. Această dificultate de reperare a tabelelor este astăzi învinsă în multe programe informatice recente, care împreună cu valoarea indicelui de omogenitate dau şi probabilitatea P.

După testul Hartley, există o diferență semnificativă între mai multe variante cu cât raportul între varianța majoră $s^2_{\max}$ şi varianța minoră $s^2_{\min}$:

$$F_{\max(p;n-1)} = \frac{s^2_{\max}}{s^2_{\min}}$$

întrece valoarea critică dată în tabelele corespunzătoare. Indicii valorii lui F_{max} consideră numărul **p** de grupe în comparație simultană şi numărul **gdl** de **n-1**, a fiecărei grupe. Testul cere ca grupele să aibă toate acelaşi număr **n** de observații. Este un test simplu dar nu robust; presupunerea fundamentală este ca datele să fie distribuite normal. Dacă nu este posibil să se presupună normalitatea distribuției pentru fiecare grupă, ar trebui să se recurgă la alte teste, ca acelea neparametrice. Aceasta dacă:

- nu există teste parametrice adaptate verificării omogenității varianței,
- când distribuția datelor se abat de la normalitate.

B. Şi testul propus de Cochran în 1967 poate să fie aplicat numai la experimente echilibrate.

Este metodologic simplu ca şi precedentul şi permite o verificare rapidă a ipotezei nule de omogenitate a diferitelor eşantioane.

Metodele lui Cochran sunt două:

- testul varianței maxime;
- testul varianței minime.

Testul varianței maxime este cel acela propus original de Cochran.

Este bazat pe raportul în varianța maximă şi suma tuturor celorlalte variante. Se calculează raportul,

$$R_{n,p} = \frac{s^2_{\max}}{s_1^2 + s_2^2 + \ldots + s_p^2} = \frac{s^2_{\max}}{\sum_{i=1}^{p} s_i^2}$$

Material cu drept de autor

unde, s^2_{max} este varianța majoră dintre eşantioanele considerate; $s_1^2, s_2^2, ..., s_p^2$ sunt variantele celor **p** grupe, cu un număr **n** de replici egale în orice grupă.

Şi în acest caz, limitele derivă din cerința unui număr egal de observații în toate grupele şi din difuzia redusă a tabelelor specifice. Cu un număr de observații foarte mare (infinit) raportul tinde către 1/t.

Testul variantei minime este dat de raportul:

$$S_{n,p} = \frac{s^2_{min}}{s_1^2 + s_2^2 + ... + s_p^2} = \frac{s^2_{min}}{\sum_{i=1}^{p} s_i^2}$$

unde, s^2_{min} este varianța minoră dintre eşantioanele considerate; $s_1^2, s_2^2, ..., s_p^2$ sunt variantele celor **p** grupe, cu un număr **n** de replici egale în orice grupă. Validitatea şi limitele sunt aceleaşi ca în testul variantei maxime.

Primul test este de utilizat când se pune ipoteza că o varianță este net mai mare decât toate celelalte, în timp ce a doua se utilizează în condiții experimentale opuse.

Tabelul 6.16

Valori critice $R_{(n,p)}$ pentru criteriul lui Cochran: n – numărul de observații într-un eşantion (grupă); p – numărul de eşantioane (grupe);

$\alpha = 0,05$

n / p	2	3	4	5	6	7	8	9	10	∞
2	0,9985	0,9750	0,9392	0,9057	0,8772	0,8534	0,8332	0,8159	0,8010	0,5000
3	0,9669	0,8709	0,7977	0,7457	0,7071	0,6771	0,6530	0,6333	0,6167	0,3333
4	0,9065	0,7679	0,6841	0,6287	0,5895	0,5598	0,5365	0,5175	0,5017	0,2500
5	0,8412	0,6838	0,5981	0,5441	0,5065	0,4783	0,4564	0,4387	0,4241	0,2000
6	0,7808	0,6161	0,5321	0,4803	0,4447	0,4184	0,3980	0,3817	0,3682	0,1667
7	0,7271	0,5612	0,4800	0,4307	0,3974	0,3726	0,3535	0,3384	0,3259	0,1429
8	0,6798	0,5157	0,4377	0,3910	0,3595	0,3362	0,3185	0,3043	0,2926	0,1250
9	0,6385	0,4775	0,4027	0,3584	0,3286	0,3067	0,2901	0,2768	0,2659	0,1111
10	0,6020	0,4450	0,3733	0,3311	0,3029	0,2823	0,2666	0,2541	0,2439	0,1000

C. Testul lui Barlett. Mai complexă este metodologia pentru **testul de semnificație a lui Barlett**.

În literatura statistică este cel mai răspândit şi oferă două avantaje față de cele două teste precedente:

Material cu drept de autor

- eşantioanele de comparat pot conține un număr diferit de replici;
- pentru verificarea semnificației între **p** grupe se utilizează distribuția $\chi^2_{(p-1)}$ cu **p-1 gdl**, mai uşor reperabilă decât distribuțiile specifice precedente a lui Hartley şi Cochran

Cu **p** măsuri a variației a probelor $\mathbf{s^2}$ care au $\mathbf{n_i}$ **gdl**, eventual diferite între ele, extrase întâmplător din populații distribuite normal, testul aproximat a lui Bartlett urmează o distribuție $\chi^2_{(p-1)}$ bazat pe raportul:

$$\chi^2_{(p-1)} = \frac{M}{C}$$

ude: C este factor de corecție propus succesiv pentru utilizarea distribuției $\chi^2_{(p-1)}$ şi este egal cu:

$$C = 1 + \frac{1}{3.(p-1)} \times \left(\sum \frac{1}{v_i} - \frac{1}{\Sigma v_i} \right)$$

Şi rezultă o valoare aproape de 1.

- M este egal cu:

$$M = \Sigma v_i . \ln \bar{s}^2 - \Sigma v_i . \ln s_i^2 ,$$

$\bar{s}^2$ este media ponderată a variațelor determinată cu relația:

$$\bar{s}^2 = \sum \frac{v_i . s_i^2}{\Sigma v_i}$$

Pentru calculul lui **M** (în unele teste este însemnat cu **B**) diverşi autori propun utilizarea logaritmului zecimal, dar este preferabil logaritmului natural; adică un alt mod pentru calculul lui M este:

$$M = 2{,}30259[(\log \bar{s}^2).(\Sigma v_i) - \Sigma v_i \log s_i^2]$$

Tabelul 6.17

Zone			
I	II	III	IV
190	138	173	198
210	149	164	207
205	128	185	232
208	136	179	184
206	152	188	193

Acest test de omogeneitate în trecut era considerat foarte puternic dar numai când distribuția datelor este normală.

Dacă distribuția datelor este **platicurtică**, valoarea probabilității α calculate este mai mare decât acea reală (testul e conservativ mai puțin potent; devine mai greu de exclus ipoteza nulă şi deci e mai uşor să se comită o eroare de speța II-a), dacă distribuția este **leptocurtică**, valoarea probabilității α calculată e mai mică decât cea reală, ruinând conceptele şi concluziile precedente.

Material cu drept de autor

Testul poate să fie aplicat pe probe nu excesiv de mici, pentru care se cere ca orice variantă să fie calculată pe o probă cu cel puțin 5, 6 observații.

Exemplu de calcul 6.5. (pentru testele Hartley, Cochran şi Bartlett cu aceleaşi date de observație). Pentru a verifica existența diferențelor în calitatea aerului, în 4 zone a unui oraş s-a măsurat cantitatea de solvenți aromatici în suspensie, tabelul 6.17. Se cere să se verifice pe baza datelor măsurate şi prezentate în patru grupe dacă datele obținute pot fi considerate omogene.

Tabelul 6.18

Variante			
I	II	III	IV
63,20	96,81	92,70	335,70

Rezolvare. Se determină variantele pentru cele patru grupe fiind prezentate în tabelul 6.18.

Amintim că fiecare din ele au 4 grade de libertate. În aceste condiții, ipoteza nulă este:

$H_0 : \sigma_I^2 = \sigma_{II}^2 = \sigma_{III}^2 = \sigma_{IV}^2$, ipoteza alternativă:

H_1 este ca cel puțin una din variante să fie diferită.

A. Metoda Hartley. Se calculează raportul între varianța majoră (335,7) şi varianța minoră (63,20) şi: se obține un F cu indici 4 (numărul de grupe) şi 4 (numărul de observații în fiecare grup mai puțin unu - **gdl**).

$$F_{(4,4)} = \frac{335,70}{63,20} = 5,30$$

Pentru semnificație se compară valoarea calculată (**5,30**) cu valoarea tabelată la probabilitatea fixată pentru numărul de grupe 4 şi **gdl** 4: pentru α **= 0,05**, rezultă **20,6** (tabelul 6.15).

B. Metoda Cochran. Se estimează un raport R:

$$R_{5,4} = \frac{335,70}{63,2 + 96,81 + 92,7 + 335,7} = 0,57$$

În tabele, valoarea critică la probabilitatea α **= 0,05**, rezultă = **0,6287**. Valoarea calculată **0,57** este inferioară celei ultime **0,6287**, deci nu este demonstrată o diferență semnificativă între cele 4 variante.

C. Metoda Barlett. Mai întâi trebuie calculată varianța medie:

$$\bar{s}^2 = \sum \frac{\nu_i . s_i^2}{\Sigma \nu_i} = \frac{4.63,2 + 4.96,81 + 4.92,7 + 4.335,7}{16} = \frac{2353,64}{16} = 147,1$$

Succesiv se estimează,

- valoarea lui M:

$$M = \Sigma \nu_i . \ln \bar{s}^2 - \Sigma \nu_i . \ln_i^2 = 16.\ln 147,1 - (4.\ln 63,2 + 4.\ln 96,81 + 4.\ln 92,7 + 4.\ln 335,7) =$$
$$= 16.4,991 - (4.4,146 + 4.4,573 + 4.4,529 + 4.5,816) = 79,856 - (16,584 + 18,292 + 18,116 +$$
$$+ 23,264) = 79,856 - 76,256 = 3,60$$

- valoarea lui C:

Material cu drept de autor

$$C = 1 + \frac{1}{3.(p-1)} \times \left(\sum \frac{1}{v_i} - \frac{1}{\Sigma v_i} \right) = 1 + \frac{1}{3.3} \times \left(\frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} \right) = 1 + \frac{1}{9} \times \left(\frac{4}{4} - \frac{1}{16} \right) =$$

$$= 1 + 0{,}111 \times (1 - 0{,}0625) = 1 + (0{,}111 \times 0{,}9375) = 1 + 0{,}104 = 1{,}104$$

Valoarea lui $\chi^2_{(3)}$, cu 3 **gdl**:

$$\chi^2_{(3)} = \frac{M}{C} = \frac{3{,}60}{1{,}104} = 3{,}26$$

În tabela valorilor critice, la probabilitatea α=**0,05**, valoarea tabelată este 7,81. Valoarea calculată 3,26 este inferioară: nu se poate exclude ipoteza nulă. După rezultatul testului, cele 4 grupe de comparat au variante nesemnificativ diferite.

Dar ce valoare are un test asupra omogenității variantei ? Pentru a accepta concluziile obținute prin intermediul analizei, rămânând aproximările deja evidențiate asupra puterii acestor teste şi care derivă din numărul redus de observări pe probă, ca în cazul exemplului utilizat.

6.6. Analiza variantei în populațiile divizate în grupe

În populațiile împărțite pe grupe se pot calcula:

- media pentru fiecare grupă şi media colectivității totale;
- varianța pentru fiecare grupă si varianța pentru întreaga colectivitate.

Factorii care determină varianța în astfel de distribuții sunt:

- factori *esențiali* (se mai numesc *înregistrați),* în funcție de care s-a realizat gruparea şi care explică abaterile mediilor de grupă de la media generală;
- factori *neesențiali* (se mai numesc *întâmplători, neînregistrați, reziduali)* adică toți ceilalți factori, în afara celor de grupare, care determină variația şi care acționează în interiorul fiecărei grupe fiind cauza abaterilor termenilor individuali de la media grupei din care fac parte.

Atunci când factorul esențial este determinant, varianța dintre grupe est mai mare decât varianța din interiorul grupelor. Exemple:

A) Studenții Facultății de Ştiințe Economice, anul doi, sunt împărțiți în şapte grupe. Dacă ne propunem o analiză a situației la învățătură pentru anul întâi de studiu vom calcula media de absolvire a anului întâi pentru fiecare grupă şi media pentru întregul an. De asemenea, vom calcula variația în fiecare grupă, variația în întregul an de studiu precum şi varianța, diferențierea, între grupe.

B) Salariații unei firme se împart în grupe după anii de vechime iar în cadrul fiecărei grupe astfel formate se reîmpart în grupe după nivelul salariului. Pentru a analiza situația vom calcula salariul mediu şi varianța pentru toți salariații, salariul mediu şi varianța pentru fiecare grupă

Material cu drept de autor

de vechime precum şi varianța, împrăştierea, între grupele formate după vechimea în muncă.

C) Pentru o analiză a nivelului impozitelor încasate într-un an într-o economie se folosesc datele unei distribuții de forma celei din tabelul următor, tabelul 1.

Tabelul 1

Grupe contribuabili după volumu impozitulu (u.m.) / Grupe contribuabili pe județe	$y_1, \ldots, y_j, \ldots, y_m$	Total unități pe grupe
x_1 . x_i . x_n	$f_{11}, \ldots, f_{1j}, \ldots, f_{1m}$ ..... $f_{i1}, \ldots, f_{ij}, \ldots, f_{im}$ ..... $f_{n1}, \ldots, f_{nj}, \ldots, f_{nm}$	f_1 . f_i . f_n
Total unități pe subgrupe (f_j)	$f_1, \ldots, f_j, \ldots, f_m$	$\sum_{i=1}^{n}\sum_{j=1}^{m} f_{ij}$

Pentru analiză *vom* calcula impozitul mediu pe un contribuabil şi varianța în rândul contribuabililor din întreaga țară, impozitul mediu pe un contribuabil şi varianța în rândul contribuabililor din fiecare județ precum şi varianța, diferențierea, dintre județe.

Analiza variației în populațiile împărțite în grupe porneşte de la o repartiție bidimensională de frecvențe rezultată în urma unei grupări după două variabile, **x** (variabila factorială, cauză) şi **y** (variabila rezultativă, efect) ca şi în exemplul de mai sus. Completând Tabelul 1 cu o coloană pentru calculul mediilor şi o coloană pentru calculul dispersiilor, aferente variabilei efect, obținem informația sistematizată ca în Tabelul 2.

Variația totală înregistrată în colectivitate după caracteristica efect poate fi analizată în următoarele sensuri:

- abaterile valorilor individuale dintr-o grupă de la media grupei respective $(y_j - \bar{y}_i)$.

- abaterile mediilor grupei de la media colectivității totale $(\bar{y}_i - \bar{y})$.

- abaterile tuturor valorilor individuale de la media valorilor colectivității $(y_j - \bar{y})$.

Între cele trei componente există următoarea relație:

$$(y_j - \bar{y}) = (\bar{y}_j - \bar{y}_i) + (\bar{y}_i - \bar{y})$$

Material cu drept de autor

care se verifică la nivelul fiecărei unități statistice din populație.

Tabelul 2

Grupe contri-buabili după y / **Grupe după x**	$y_1,\ldots,y_j,\ldots,y_m$	Total unități pe grupe	Medii de grupă, $\bar{y}_i$	Dispersii de grupă σ_i^2
x_1	$f_{11},\ldots,f_{1j},\ldots,f_{1m}$	f_1	$\bar{y}_1$	σ_1^2
.	.....	.	.	.
x_i	$f_{i1},\ldots,f_{ij},\ldots,f_{im}$	f_i	$\bar{y}_i$	σ_i^2
.	.....	.	.	.
x_n	$f_{n1},\ldots,f_{nj},\ldots,f_{nm}$	f_n	$\bar{y}_n$	σ_n^2
Total unități pe subgrupe (f_j)	$f_1,\ldots,f_j,\ldots,f_m$	$\sum_{i=1}^{n}\sum_{j=1}^{m} f_{ij}$	$\bar{y}$	σ_y^2

Indicatorii de variație prin care se caracterizează cele trei abateri definite mai sus sunt:

Dispersia totală, calculată pe baza tuturor abaterilor individuale față de media colectivității totale.

$$\sigma^2 = \frac{\sum_{j=1}^{m}(y_j - \bar{y})^2 f_j}{\sum_{j=1}^{m} f_j}$$

unde: $\mathbf{y_j}$ - valorile caracteristicii distribuită în funcție de factorul de grupare **x**; $\bar{y}$ - media caracteristicii rezultative pentru întreaga colectivitate; $\mathbf{f_j}$ -frecvențele subgrupelor formate după variația caracteristicii rezultative; **m** - numărul subgrupelor.

Media colectivității totale s-a calculat:

$$\bar{y} = \frac{\sum_{i=1}^{n}\bar{y}_i \cdot f_i}{\sum_{i=1}^{n} f_i}$$

Material cu drept de autor

unde: $\bar{y}_i$ - mediile pe grupe (medii condiționate); $\mathbf{f_i}$ - volumul grupelor formate după variația caracteristicii de grupare; n - numărul grupelor.

Acest indicator reuneşte influențele tuturor factorilor, esențiali şi neesențiali, care determină variația caracteristicii **y**.

Dispersiile de grupe, calculată pe baza abaterilor tuturor variantelor dintr-o grupă față de media lor de grupă.

$$\sigma_i^2 = \frac{\sum_{j=1}^{m}(y_j - \bar{y}_i).f_{ij}}{\sum_{j=1}^{m} f_{ij}}$$

unde: $\mathbf{y_j}$ - valorile caracteristicii distribuită în funcție de factorul de grupare **x**; $\bar{y}_i$ - mediile pe grupe (medii condiționate); $\mathbf{f_{ij}}$ - frecvențele condiționate de variația condiționată a caracteristicilor **x** şi **y**.

Mediile pe grupe s-au calculat:

$$\bar{y}_i = \frac{\sum_{j=1}^{m} y_j \cdot f_{ij}}{\sum_{j-1} f_{ij}}, \qquad i = 1,2,\ldots,n$$

Acest indicator sintetizează influența factorilor aleatori care acționează în interiorul grupelor determinând variația valorilor individuale din acestea.

Media dispersiilor de grupă, calculată ca o medie aritmetică a dispersiilor tuturor grupelor.

$$\bar{\sigma}^2 = \frac{\sum_{i=1}^{n} \sigma_i^2 \cdot f_i}{\sum_{i=1}^{n} f_i}$$

Acest indicator măsoară influența factorului de grupare după o relație de directă proporționalitate.

Mediile şi dispersiile de grupă se mai numesc medii şi dispersii *condiționate*.

Media şi dispersia pe întreaga colectivitate se numesc *marginale*.

Regula adunării dispersiilor. Dispersia colectivității totale este egală cu media dispersiilor de grupă plus dispersia dintre grupe.

$$\sigma^2 = \bar{\sigma}^2 + \sigma_{y/x}^2$$

Material cu drept de autor

Pe baza acestei reguli se calculează indicatorii.

Coeficientul de determinație, care măsoară influența factorului de grupare **x** asupra variației caracteristicii **y**.

$$R^2 = \frac{\sigma^2_{y/x}}{\sigma^2} 100$$

Se consideră că factorul de grupare **x** influențează hotărâtor variația caracteristicii rezultative **y** dacă $\mathbf{R^2 > 50\%}$.

Coeficientul de nedeterminație, care măsoară influența factorilor neesențiali asupra variației caracteristicii y.

Cei doi indicatori sunt complementari $\mathbf{R^2 + N^2 = 100\%}$.

Exemplu de calcul 6.6. Se consideră distribuția unui eşantion de studenți după sex şi vârstă, astfel:

Tabelul 3

Grupe de studenți, după sex	Subgrupe de studenți după vârstă (ani împliniți)					Total
	18-20	20-22	22-24	24-26	26-28	
Feminin	4	14	14	8	-	40
Masculin	2	2	6	6	4	20
Total subgrupă	6	16	20	14	4	60
Observație. Date convenționale, limita inferioară inclusă în interval.						

Să se analizeze varianța în distribuția dată.

Rezolvare. Caracteristica de grupare **x** se consideră caracteristica sex, iar caracteristica secundară **y** caracteristica vârstă.

Pentru calcule şi analiză departajăm:

Tabelul 4, Grupa 1-a, feminin

Grupe stud. după vârstă	Număr studenți	Centrul de interval, y_j	$y_j.f_{ij}$	$(y_j - \bar{y}_i)^2 .f$
18-20	4	19	76	43,56
20-22	14	21	294	23,66
22-24	14	23	322	6,86
24-26	8	25	200	58,32
Total	40	-	892	132,4

Material cu drept de autor

Vârsta medie a fetelor:

$$\bar{y}_1 = \frac{\sum_{j=1}^{4} y_j f_{1j}}{\sum_{j=1}^{4} f_{1j}} = \frac{892}{40} = 22{,}3 \; ani$$

Dispersia grupei 1:

$$\sigma_1^2 = \frac{\sum_{j=1}^{4} (y_j - \bar{y}_1)^2 . f_{1j}}{\sum_{j=1}^{4} f_{1j}} = \frac{132{,}4}{40} = 3{,}31$$

Abaterea medie pătratică pentru grupa 1:

$$\sigma_1 = \sqrt{\sigma_1^2} = \sqrt{3{,}31} = 1{,}82$$

Coeficientul de variație pentru grupa 1:

$$v_1 = \frac{\sigma_1}{\bar{y}_1} 100 = \frac{1{,}82}{22{,}3} 100 = 8{,}16\%$$

Afirmație. Grupa de sex feminin este foarte omogenă din punct de vedere al vârstei, iar media de 22,3 ani este reprezentativă.

Tabelul 5 Grupa 2-a, masculin

Grupe stud. după vârstă	Număr studenți	Centrul de interval, y_j	$y_j.f_{ij}$	$(y_j - \bar{y}_i)^2 . f$
18-20	2	19	38	46,08
20-22	2	21	42	15,68
22-24	6	23	138	3,84
24-26	6	25	150	8,64
26-28	4	27	108	40,96
Total	40	-	476	115,2

Vârsta medie a băieților:

Material cu drept de autor

$$\bar{y}_2 = \frac{\sum_{j=1}^{5} y_j f_{2j}}{\sum_{j=1}^{5} f_{2j}} = \frac{476}{20} = 23,8 \ ani$$

Dispersia grupei 2:

$$\sigma_2^2 = \frac{\sum_{j=1}^{5} (y_j - \bar{y}_2)^2 \cdot f_{1j}}{\sum_{j=1}^{5} f_{2j}} = \frac{115,2}{20} = 5,76$$

Abaterea medie pătratică pentru grupa 2:

$$\sigma_2 = \sqrt{\sigma_2^2} = \sqrt{5,76} = 2,4$$

Coeficientul de variație pentru grupa 2:

$$v_2 = \frac{\sigma_2}{\bar{y}_2} 100 = \frac{2,4}{23,8} 100 = 10,08\%$$

Afirmație. Grupa de sex masculin este, de asemenea, foarte omogenă din punct de vedere al vârstei, dar mai puțin omogenă decât grupa de sex feminin, iar media de 23,8 ani este reprezentativă.

Tabelul 6, Total eşantion

Grupe stud. după vârstă	Număr studenți	Centrul de interval, y_j	$y_j.f_{ij}$	$(y_j - \bar{y}_i)^2 \cdot f_{ij}$
18-20	6	19	114	86,64
20-22	16	21	336	51,84
22-24	20	23	460	0,8
24-26	14	25	350	67,76
26-28	4	27	108	70,56
Total	60	-	1368	277,6

Vârsta medie pentru întregul eşantion:

$$\bar{y} = \frac{\sum_{j=1}^{5} y_j f_j}{\sum_{j=1}^{5} f_j} = \frac{1368}{60} = 22,8 \ ani$$

Material cu drept de autor

Dispersia eşantionului:

$$\sigma^2 = \frac{\sum_{j=1}^{5}(y_j - \bar{y})^2 . f_j}{\sum_{j=1}^{5} f_j} = \frac{277,6}{60} = 4,63$$

Abaterea medie pătratică pentru eşantion:

$$\sigma = \sqrt{\sigma^2} = \sqrt{4,63} = 2,15$$

Coeficientul de variație pentru eşantion:

$$v = \frac{\sigma}{\bar{y}} 100 = \frac{2,15}{22,8} 100 = 9,43\%$$

Afirmație. Eşantionul, ca şi cele două grupe din care se compune, este foarte omogen din punct de vedere al vârstei, dar mai puțin omogenă decât grupa de sex feminin, iar media de 22,8 ani este reprezentativă.

Se calculează media dispersiilor de grupă:

$$\bar{\sigma}^2 = \frac{\sum_{i=1}^{2} \sigma_i^2 . f_i}{\sum_{i=1}^{2} f_i} = \frac{3,31.40 + 5,76.20}{60} = 4,1266$$

Se calculează dispersia dintre grupe:

$$\sigma_{y/x}^2 = \frac{\sum_{i=1}^{2}(\bar{y}_i - \bar{y})^2 . f_i}{\sum_{i=1}^{2} f_i} = \frac{(22,3 - 22,8)^2 .40 + (23,8 - 22,8)^2 .20}{60} = 0,5$$

Se verifică regula adunării dispersiilor:

$$\sigma^2 = \bar{\sigma}^2 + \sigma_{y/x}^2 = 0,5 + 4,13 = 4,63$$

Se calculează coeficientul de determinație:

$$R^2 = \frac{\sigma_{y/x}^2}{\sigma^2} 100 = \frac{0,5}{4,63} 100 = 11\%$$

Se calculează coeficientul de nedeterminație:

$$N^2 = \frac{\bar{\sigma}^2}{\sigma^2} 100 = \frac{4,13}{4,63} 100 = 89\%$$

Se verifică complementaritatea coeficienților:

$$11\% + 89\% = 100\%$$

Material cu drept de autor

Afirmație. Întrucât coeficientul de determinație $\mathbf{R^2}$ este mult sub 50%, înseamnă că variabila sex nu este determinantă pentru variabila vârstă, aceasta din urmă fiind influențată de alți factori sau fiind factor de influență pentru alte variabile.

Exemplul de calcul 6.7. Într-o colectivitate statistică s-au cules date privitoare la două variabile numerice, obținându-se:

$$\{x_i\} = \{4,1,1,5,6,3,2,1\} \text{ şi}$$

$$\{y_i\} = \{100,90,40,80,70,50,100,700\}$$

Se cere să se arate după care din variabile colectivitatea este mai omogenă.

Rezolvare.

Pentru variabile **X** se va calcula:

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{23}{8} = 2,875$$

$$s_x^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1} = \frac{26,875}{7} = 3,839$$

Sau, prin metoda momentelor:

$$s_x^2 = \frac{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}}{n-1} = \frac{93 - \frac{(23)^2}{8}}{7} = 3,839$$

$$s_x = \sqrt{s_x^2} = \sqrt{3,839} = 1,959; \quad Cv = \frac{s_x}{\bar{x}}.100 = \frac{1,959}{2,875}.100 = 68,1\%$$

Pentru variabila Y vom avea:

$$\bar{y} = \frac{\Sigma y_i}{n} = \frac{600}{8} = 75$$

$$s_y^2 = \frac{\Sigma(y_i - \bar{y})^2}{n-1} = \frac{3400}{7} = 485,71$$

$$s_y = \sqrt{s_y^2} = \sqrt{485,71} = 22,039; \quad Cv = \frac{s_y}{\bar{y}}.100 = \frac{22,039}{75}.100 = 29,38\%$$

Rezultă că seria după caracteristica Y este mai omogenă decât cea formată după caracteristica X.

Material cu drept de autor

Cap.7. CORELAȚIE ŞI REGRESIE

7.1. Considerații preliminare

De la început, când omul şi-a pus primele întrebări asupra diverselor fenomene din natură s-a întrebat: care ar fi cauza producerii lor ? A descoperit astfel principiul cauzalității care acționează în toate împrejurările, după care a stabilit şi principiul determinismului manifestat în legile naturii. **A descoperit astfel că în natură şi viața social-economică există fenomene „cauză" şi fenomene „efect"**.

Transpusă această situație în limbaj probabilistico-statistic se poate spune că se manifestă în eşantioane de cercetări statistice existența de variabile independente (cauză) şi variabile dependente (efect).

În practica cercetării există situați de existență a unei variabile sau mai multe variabile independente şi una sau mai multe variabile dependente.

Pentru situații de acest gen, problema care se pune este de a evalua cantitativ existența unei relații între variația reciprocă a acelor două categorii de variabile. Testul statistic utilizat este testul de corelație (**coeficientul de corelație**). Termenul de corelație, înainte de a fi un concept statistic este un cuvânt uzual în limbajul cotidian. În esență, el exprimă o legătură între anumite aspecte ale realității aşa cum este ea reflectată în planul observației directe (o parcare plină cu maşini ne sugerează că magazinul alăturat este plin cu cumpărători, între numărul de maşini din parcare şi numărul de cumpărători existând o anumită „corelare").

La nivel statistic, corelația exprimă o legătură cantitativă sistematică între valorile a două variabile perechi, măsurate pe subiecți aparținând aceluiaşi eşantion de cercetare.

7.2. Forme de legături existente între fenomene şi procese economico-sociale

Fenomenele şi procesele economico-sociale au un caracter complex, manifestat, deseori, printr-o legătură de cauzalitate între mărimile de intrare şi de ieşire din sistemul format de aceste fenomene şi procese economico-sociale. Există diferite forme de legături, funcționale şi statistice, natura acestora stabilindu-se printr-o analiză calitativă multilaterală.

Fie două variabile şi anume, variabila independentă **x** şi variabila **y** care este dependentă de prima.

Dacă pentru o valoare determinată a variabilei independente **x**, variabila dependentă **y** ia, de asemenea, o valoare determinată, discretă, se spune că între cele două variabile există o **legătură funcțională: y = f(x)**.

Dacă fiecărei valori **x** îi corespunde nu o singură valoare **y** ci o **repartiție legată** de valori **y** (legată de valoarea lui **x**), caracterizată de fiecare **x** prin câte o medie $\overline{y}$ legată, avem de-a face cu o legătură

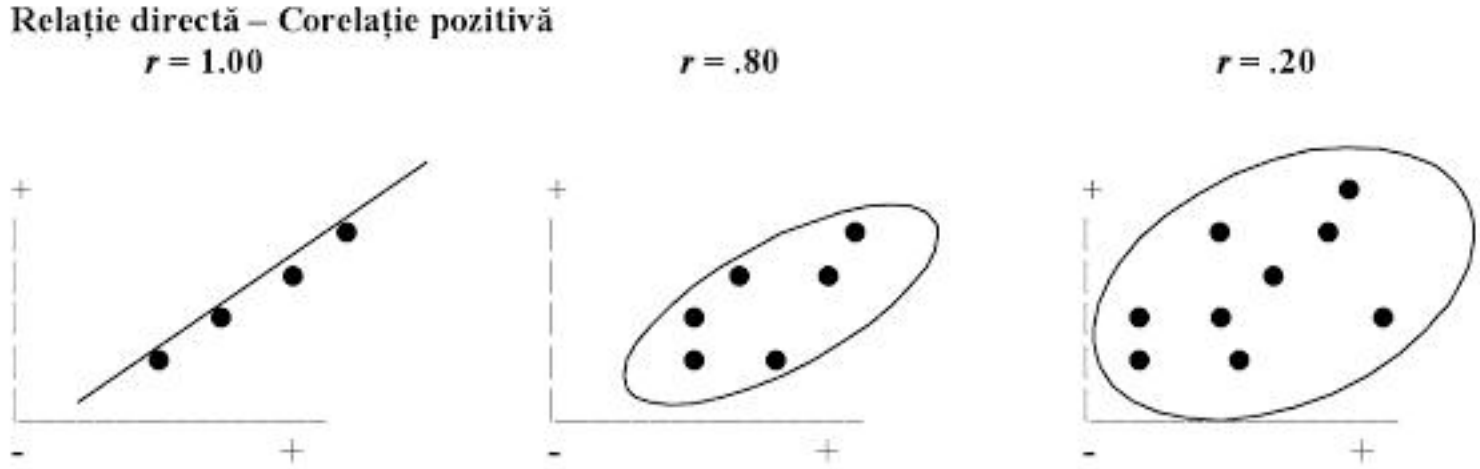
Material cu drept de autor

statistică (stohastică), respectiv există o **corelație între y şi x**. Dacă fixăm valoarea lui **x**, putem estima valoarea repartiției legate **y**. Dacă mediile repartițiilor legate se plasează aproximativ de-a lungul unei drepte, dreapta care trece cel mai aproape de poziția centrală a punctelor se numeşte dreaptă de regresie a lui y în raport cu x, iar corelația se numeşte corelație simplă (liniară).

Reprezentarea grafică a legăturilor statistice. Plasarea valorilor celor două variabile pe un grafic produce o imagine intuitivă a relației dintre valori. Acest tip de grafic se numeşte *scatterplot*.

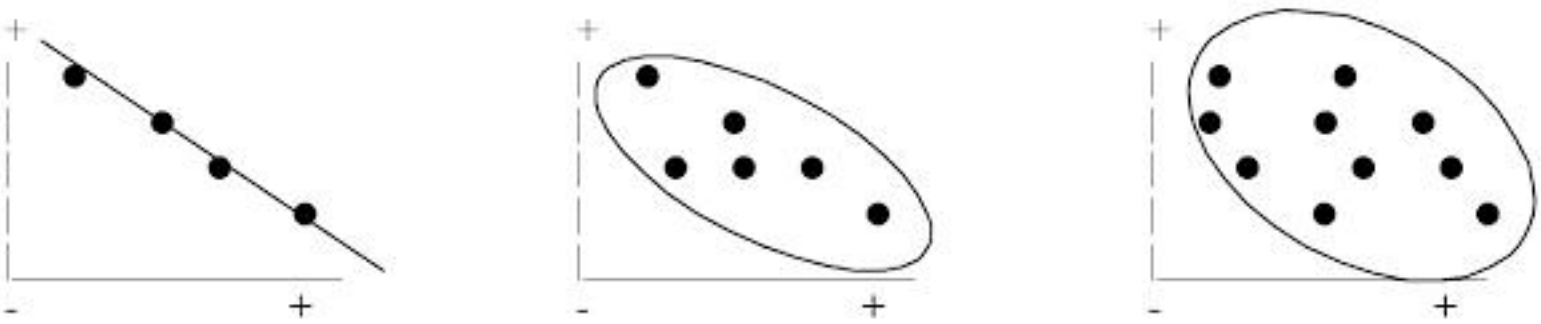
În cazul unei corelații pozitive, reprezentările *scatterplot* pot arăta astfel:

Relație directă – Corelație pozitivă



În cazul unei corelații pozitive tendința este aceea ca valorilor mari de pe axa orizontală să le corespundă valori mari pe axa verticală. În cazul unei corelații pozitive perfecte (**r=+1**), punctele de intersecție ale perechilor de valori se plasează pe o linie. Cu cât corelația este mai mică, cu atât norul de puncte este mai larg dar forma elipsei indică relația pozitivă dintre cele două variabile.

Relație indirectă- Corelație negativă



În cazul unei corelații negative tendința este aceea ca valorilor mari de pe axa orizontală să le corespundă valori mici pe axa verticală. Ca urmare, atât linia corelației negative perfecte (**r = -1**) cât şi diagonala mare a elipsei norului de puncte al corelației imperfecte se orientează din stânga sus spre dreapta jos a sistemului de coordonate.

Atunci când corelația dintre cele două variabile este inexistentă, norul punctelor de intersecție are o formă circulară, care nu conturează nici o tendință (**r = 0**). În imaginea de mai jos avem reprezentări *scatterplot* caracteristice pentru corelații liniare negative.

Material cu drept de autor

Estimația poziției dreptei de regresie care exprimă legătura între variabilele independente şi variabilele dependente are o precizie cu atât mai mare cu cât corelația este mai intensă.

Pentru analiza existenței şi intensității legăturilor statistice s-au elaborat mai multe metode, cele mai uzuale fiind: ***metoda corelației; metoda grafică; metoda tabelară (a tabelului de corelație şi a tabelului de asociere); metoda analizei dispersionale***.

7.3. Covarianța şi corelația

Gradul de legătură între variabilele **x** şi **y** este măsurată prin momentul centrat mixt de ordinul doi numit covarianță:

$$s_{XY} = m_{1/1} = \frac{\Sigma(x-\bar{x})\times(y-\bar{y})}{n-1} \qquad (7.1)$$

în care **n** este numărul de observații perechi (**x,y**), iar $\overline{x}$ şi $\overline{y}$ sunt valorile medii pentru variabilele respective.

Pentru variabile aleatoare continue momentul centrat mixt se determină cu expresia:

$$\sigma_{XY} = \int_{-\infty-\infty}^{+\infty+\infty}\int (x-m_x).(y-m_y).f(x,y).dx.dy \qquad (7.2)$$

în care $\mathbf{m_x}$ şi $\mathbf{m_y}$ sunt valorile medii ale variabilelor aleatoare **X** şi **Y**, iar **f(x,y)** este funcția bidimensională de dependență a variabilelor **X** şi **Y**.

Expresia dată de relația (7.1), aşa cum se observă, este covarianța estimată pe baza unei selecții. Pentru întreaga populație statistică, covarianța se notează σ_{xy} şi se determină cu expresia (7.2).

7.4. Coeficientul de corelație

Coeficientul teoretic de corelație $\rho_{Y/X}$, a două variabile aleatoare **X** şi **Y,** este covarianța variabilelor aleatoare normate corespunzătoare:

$$\rho_{Y/X} = Cov\left(\frac{X-\overline{X}}{\sigma_X}, \frac{Y-\overline{Y}}{\sigma_Y}\right) = M\left[\left(\frac{X-\overline{X}}{\sigma_X}\right)\left(\frac{Y-\overline{Y}}{\sigma_Y}\right)\right] = \qquad (7.3)$$

$$= \frac{M\left[\left(X-\overline{X}\right)\left(Y-\overline{Y}\right)\right]}{\sigma_X\sigma_Y} = \frac{Cov(XY)}{\sigma_X\sigma_Y} = \frac{M_{Y/X}}{\sigma_X\sigma_Y}$$

Se demonstrează că valorile extreme pe care le poate lua coeficientul de corelație sunt **+1** şi **-1.**

Dacă variabilele **X** şi **Y** sunt independente:

Cov(X,Y) = 0, $\rho_{Y/X} = 0$ (7.4)

Reciproca însă nefiind adevărată.

Material cu drept de autor

7.5. Corelație simplă (liniară)

În cazul seriilor empirice de repartiție simplă:

$$r_{y/x} = r = \frac{\operatorname{cov}(x,y)}{\sigma_x \sigma_y} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{(n-1)\sigma_x \sigma_y} = \frac{s_{y/x}}{s_x s_y} \tag{7.5}$$

- în cazul unui număr mare de date sistematizate prin:

a. grupare simplă:

$$r_{y/x} = r = \frac{\operatorname{cov}(x,y)}{\sigma_x \sigma_y} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{(\Sigma n_i)\sigma_x \sigma_y} \tag{7.6}$$

b. grupare combinată după ambele variabile $\mathbf{X_i, Y_i}$:

$$r = \frac{\Sigma n_i \Sigma x_i y_i n_i - \Sigma x_i n_i \Sigma y_i n_i}{\sqrt{[\Sigma n_i \Sigma x_i^2 n_i - (\Sigma x_i n_i)^2][\Sigma n_i \Sigma y_i^2 n_i - (\Sigma y_i n_i)^2]}} \tag{7.7}$$

Dacă se foloseşte tabelul de corelație, în care perechile $\mathbf{x_i y_i}$ se întâlnesc de $\mathbf{n_{ij}}$ ori:

$$r = \frac{\Sigma\Sigma n_{ij} \Sigma x_i y_j n_{ij} - \Sigma x_i n_{i\bullet} \Sigma y_j n_{\bullet j}}{\sqrt{[\Sigma\Sigma n_{ij} \Sigma x_i^2 n_{i\bullet} - (\Sigma x_i n_{i\bullet})^2][\Sigma\Sigma n_{ij} \Sigma y_j^2 n_{\bullet j} - (\Sigma y_j n_{\bullet j})^2]}} \tag{7.8}$$

În practică, suma produselor $\Sigma(x-\bar{x})(y-\bar{y})$ se calculează mai comod dacă este adusă la forma:

$$\Sigma(x-\bar{x})(y-\bar{y}) = \Sigma xy - \frac{\Sigma x \Sigma y}{n} \tag{7.9}$$

când covarianța devine:

$$s_{Y/X} = m_{1/1} = \frac{\Sigma xy - \dfrac{\Sigma x \Sigma y}{n}}{n-1} \tag{7.10}$$

iar coeficientul de corelație:

$$r_{Y/X} = \frac{\Sigma xy - \dfrac{\Sigma x \Sigma y}{n}}{(n-1)s_x s_y} = \frac{s_{y/x}}{s_x s_y} \tag{7.11}$$

Abaterile standard $\mathbf{s_x}$ şi $\mathbf{s_y}$ se determină pe baza relațiilor de calcul ale acestora prezentate anterior.

Coeficientul de corelație poate varia, după cum s-a arătat anterior, între **-1** şi **+1**. Cu cât este mai apropiat de valoarea **-1** sau **+1**, corelația liniară este mai intensă. Dacă variabilele sunt necorelate, coeficientul de corelație este **0**. Dacă $\mathbf{r_{xy} > 0}$ se spune că corelația este pozitivă, adică la creşterea lui **x** creşte şi **y**, iar dacă $\mathbf{r_{xy} < 0}$ se spune că corelația este negativă, adică la creşterea lui **x**, valoarea lui **y** scade.

Material cu drept de autor

Momentul centrat mixt de ordinul doi pentru datele empirice finite se poate determina astfel:

$$m_{1/1} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \qquad (7.12)$$

în care $\mathbf{x_i}$ şi $\mathbf{y_i}$ sunt variabilele curente independente **x** şi dependente **y**; $\bar{x}$ şi $\bar{y}$ - valorile medii ale variabilelor respective.

Momentul centrat (7.12) este mai simplu de calculat pe baza expresiei:

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n}x_i y_i - \frac{1}{n}\sum_{i=1}^{n}x_i \sum_{i=1}^{n}y_i \qquad (7.13)$$

Dispersiile s_x^2 şi s_y^2 se determină folosind una din metodologiile prezentate anterior.

La un volum mare al datelor de observaţie (**n > 60**) calculele devin dificile. Pentru simplificarea lor, datele de observaţie se grupează pe intervale de valori fiind prezentate sub formă de tabel de corelaţie (tabelul 7.1 sau tabelul 7.2), iar momentul centrat mixt se determină cu expresia:

$$m_{1/1} = \frac{1}{n-1}\sum_{i=1}^{m}\sum_{j=1}^{k}n_{ij}(x_i - \bar{x})(y_j - \bar{y}) \qquad (7.14)$$

sau, mult mai simplu, cu relaţia:

$$m_{1/1} = \frac{1}{n-1}\left(\sum_{i=1}^{m}\sum_{j=1}^{k}n_{ij}x_i y_j - \frac{1}{n}\sum_{i=1}^{m}n_{i\bullet}x_i \sum_{j=1}^{k}n_{\bullet j}y_j\right) \qquad (7.15)$$

Tabelul 7.1

Tabel de corelaţie pentru distribuţia bidimensională a legăturilor statistice

x_i	y_j							Suma	$\bar{y}_i$
	1	2	3	...	j	...	k		
1	n_{11}	n_{12}	n_{13}	...	n_{1j}	...	n_{1k}	$n_{1\bullet}$	$\bar{y}_1$
2	n_{21}	n_{22}	n_{23}	...	n_{2j}	...	n_{2k}	$n_{2\bullet}$	$\bar{y}_2$
3	n_{31}	n_{32}	n_{33}	...	n_{3j}	...	n_{3k}	$n_{3\bullet}$	$\bar{y}_3$
...	...	...	...	...	...	...	...	...	...
i	n_{i1}	n_{i2}	n_{i3}	...	n_{ij}	...	n_{ik}	$n_{i\bullet}$	$\bar{y}_i$
...	...	...	...	...	...	...	...	...	...
m	n_{m1}	n_{m2}	n_{m3}	...	n_{mj}	...	n_{mk}	$n_{m\bullet}$	$\bar{y}_m$

Material cu drept de autor

Suma	$n_{\bullet 1}$	$n_{\bullet 2}$	$n_{\bullet 3}$	...	$n_{\bullet j}$	...	$n_{\bullet k}$	n	$\overline{y}$
$\overline{x}_j$	$\overline{x}_1$	$\overline{x}_2$	$\overline{x}_3$	...	$\overline{x}_j$	...	$\overline{x}_k$	$\overline{x}$	

În acest caz abaterile medii pătratice se determină cu expresiile:

$$s_x = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{m} n_{i\bullet} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{m} n_{i\bullet} x_i\right)^2\right]} \qquad (7.16)$$

şi respectiv:

$$s_y = \sqrt{\frac{1}{n-1}\left[k\sum_{j=1}^{m} n_{\bullet j} y_j^2 - \frac{1}{n}\left(\sum_{j=1}^{k} n_{\bullet j} y_j\right)^2\right]} \qquad (7.17)$$

În expresiile (7.14)...(7.17) s-au adoptat notațiile: **m** şi **k** – numărul intervalelor corespunzătoare pentru variabilele **x** şi **y**; $\mathbf{x_i}$ şi $\mathbf{y_j}$ – valorile **x** şi **y** în interiorul intervalelor respective; $\mathbf{n_{ij}}$ – numărul observațiilor comune ale variabilelor **x** şi **y**; $n_{\bullet j}$ - numărul sumat al valorilor **x** corespunzătoare aceleiaşi valori **Y = $\mathbf{y_i}$**; $n_{i\bullet}$ - numărul sumat al valorilor **y** corespunzătoare aceleiaşi valori **X = $\mathbf{x_i}$**; **n** – numărul total al observațiilor; $\overline{x}_j$ - valoarea medie a variabilelor **x**, corespunzător valorii **Y = $\mathbf{y_i}$**; $\overline{y}_i$ - valoarea medie a variabilelor **y**, corespunzător valorii **X = $\mathbf{x_j}$**; $\overline{x}$ şi $\overline{y}$ - valorile medii pentru variabilele **X** şi **Y**.

Evaluarea valorilor medii se face cu expresiile:

$$\overline{x}_j = \frac{\sum_{i=1}^{m} n_{ij} x_i}{n_{\bullet j}} \qquad (7.18) \qquad \overline{y}_i = \frac{\sum_{i=1}^{m} n_{ij} y_j}{n_{i\bullet}} \qquad (7.19$$

$$\overline{x} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{k} n_{ij} x_i}{n} \qquad (7.20) \qquad \overline{y} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{k} n_{ij} y_j}{n} \qquad (7.21)$$

sau:

Material cu drept de autor

$$\bar{x} = \frac{\sum_{j=1}^{k} n_{\bullet j}\bar{x}_j}{n} \qquad (7.22) \qquad \bar{y} = \frac{\sum_{i=1}^{m} n_{i\bullet}\bar{y}_i}{n} \qquad (7.23)$$

Ansamblul termenilor $\mathbf{n_{1j}, n_{2j},\ldots,n_{ij},\ldots,n_{mj}}$, ai variabilei întâmplătoare **X**, pe baza termenului fixat al altei variabile întâmplătoare $\mathbf{Y = y_j}$, reprezintă seria de distribuție condițională a valorii **X** pentru $\mathbf{Y = y_j}$. În felul acesta, tabelul 7.1 conține **k** distribuții condiționale a valorii **X** pentru $\mathbf{Y = y_j}$ **(j = 1, 2, 3,...,k)** şi **m** distribuții condiționale a valorii **Y** pentru $\mathbf{X = x_i}$ **(i = 1, 2, 3,...,m)**.

Ca parametri ai distribuției condiționale servesc mediile condiționale şi dispersiile condiționale. Ca evaluare a mediilor condiționale o constituie valorile $\bar{x}_j$ şi $\bar{y}_i$, iar ca evaluare a dispersiilor condiționale – dispersiile condiționale determinate cu expresiile:

$$s^2_{x/y_j} = \frac{\sum_{i=1}^{m} n_{ij}(x_i - \bar{x}_j)^2}{n_{\bullet j} - 1} \qquad (7.24)$$

şi, respectiv:

$$s^2_{y/x_i} = \frac{\sum_{j=1}^{k} n_{ij}(y_j - \bar{y}_i)^2}{n_{i\bullet} - 1} \qquad (7.25)$$

a căror calcul este mai simplu să se efectueze cu următoarele relații:

$$s^2_{x/y_j} = \frac{1}{n_{\bullet j} - 1}\left[\sum_{i=1}^{m} n_{ij}x_i^2 - \frac{\left(\sum_{i=1}^{m} n_{ij}x_i\right)^2}{n_{\bullet j} - 1}\right] \qquad (7.26)$$

respectiv:

$$s^2_{y/x_i} = \frac{1}{n_{i\bullet} - 1}\left[\sum_{j=1}^{k} n_{ij}y_j^2 - \frac{\left(\sum_{j=1}^{k} n_{ij}y_j\right)^2}{n_{i\bullet} - 1}\right] \qquad (7.27)$$

Penultima coloană şi penultimul rând din tabelul 7.1 reprezintă distribuirea particulară a valorilor întâmplătoare **X** şi **Y**, corespunzător

Material cu drept de autor

evaluărilor, a căror parametri constituie valorile $\bar{x}$, $\bar{y}$, s_x^2 şi s_y^2 [vezi expresiile (7.16), (7.17), (7.20)...(7.23)].

Coeficientul de corelație calculat, ca şi alte caracteristici empirice, constituie o valoare întâmplătoare şi poate primi diferite valori aleatoare la repetarea experienței. La analiza variabilelor independente, pentru care coeficientul general de corelație ρ este egal cu zero, coeficientul empiric **r** poate să difere esențial de zero. În legătură cu aceste aspect se naşte o importantă problemă practică, rezolvându-se prin verificarea ipotezei de absență a corelației între valorile întâmplătoare cercetate **X** şi **Y**, adică la verificarea ipotezei nule, despre egalitatea cu zero a coeficientului general de corelație pe baza datelor experimentale.

Exemplu de calcul 7.1. Se cere să se determine coeficientul de corelație dintre rezultatele obținute de opt studenți la două teste de la statistică, tabelul 7.2.

Tabelul 7.2

x	y	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$	x^2	y^2	x.y
5	6	-2	-2	4	4	4	25	36	30
9	8	0	0	4	0	0	81	64	72
7	9	1	1	0	1	0	49	81	63
10	10	2	2	9	4	6	100	100	100
6	9	1	1	1	1	-1	36	81	54
5	6	-2	-2	4	4	4	25	36	30
6	7	-1	-1	1	1	1	36	49	42
8	9	1	1	1	1	1	64	81	72
56	64	0	0	24	16	15	416	528	463

Pe baza datelor din tabelul 7.2 rezultă:

$$\bar{x} = 56/8 = 7;\ s_x^2 = 24/8 = 3;\ s_x = \sqrt{3} = 1{,}73$$

$$\bar{y} = 64/8 = 8;\ s_y^2 = 16/8 = 2;\ s_y = \sqrt{2} = 1{,}41$$

$$r = \frac{15}{8 \times 1{,}73 \times 1{,}41} = 0{,}76;\ r = \frac{8 \times 463 - 56 \times 64}{\sqrt{(18.416 - 56^2)(8.528 - 64^2)}} = 0{,}76$$

Material cu drept de autor

7.6. Regresia liniară simplă cu o singură variabilă independentă

După stabilirea unei anumite dependențe între o variabilă aleatoare independentă **X** şi o variabilă aleatoare dependentă **Y**, la analiza de corelație, se va căuta să se precizeze modul de variație a acestei dependențe.

Se va considera cazul cel mai simplu, care este variația liniară a acestei dependențe ca în figura 7.1.

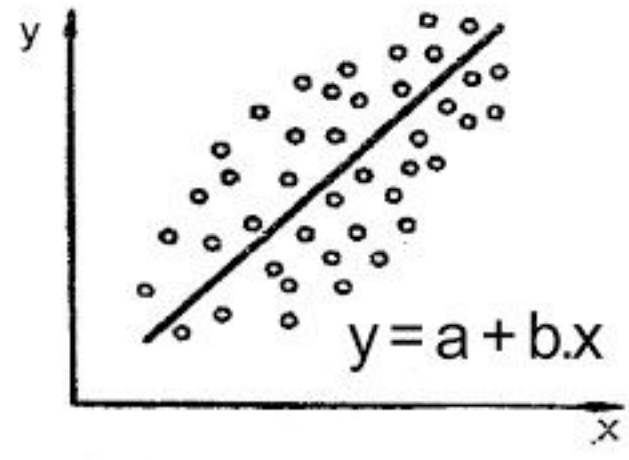


Fig.7.1. Dependența liniară, **y = f(x)**

În figura 7.1 se observă că în practică pentru o anumită valoare **X** se pot obține mai multe valori **y**, sau pentru mai multe valori **x** să rezulte din experiment aceeaşi valoare **Y**. Însă, se constată că norul de puncte, care marchează perechile x şi y, are o anumită deplasare în lungul şi în jurul unei drepte imaginare.

Problema principală care se pune la analiza regresiei liniare constă în a determina cât mai exact ecuația acestei drepte imaginare şi, deci, poziționarea ei concretă în norul de puncte ce marchează dependențele dintre **X** şi **Y**.

Majoritatea problemelor legate de analiza regresiei liniare a fenomenelor economico-sociale au la bază presupunerea cu **y** este o variabilă întâmplătoare, supuse legii normale de distribuție, iar **x** poate avea valori întâmplătoare sau neîntâmplătoare.

Dacă variabilele cercetate sunt dependente, atunci schimbarea valorii **x**, în cazul general, se pot schimba ambii parametri ai distribuției normale pentru variabila **y** şi anume: valoarea medie şi dispersia:

$$m_y = f_1(x) \tag{7.28}$$

şi respectiv:

$$\sigma_y^2 = f_2(x) \tag{7.29}$$

Prima dependență (7.28) se numeşte **linie medie de regresie**, iar a doua dependență (7.29) se numeşte **oscilație admisibilă în jurul liniei medii de regresie**.

Analiza de regresie a rezultatelor experimentale prevede evaluarea parametrilor de identificare concretă a liniei de regresie şi evidențiere a oscilației datelor experimentale în jurul acestei linii de regresie. De asemenea, determinarea intervalului de încredere pentru parametrii evaluați.

Dacă ambele valori cercetate, **X** şi **Y**, sunt întâmplătoare, atunci, în mod obişnuit, analiza de regresie precede analiza de corelație, pe baza

Material cu drept de autor

căreia se efectuează evaluarea mediei şi ai celorlalți parametrii folosiți la analiza de corelație ($\bar{x}$, $\bar{y}$, s_x^2, s_y^2, r).

Pentru evaluarea liniei medii $\mathbf{m_y}$ se foloseşte ecuația:

$$m_y = M(Y/X) = \alpha + \rho \frac{\sigma_y}{\sigma_x}(x - \bar{x}) \tag{7.30}$$

în care **M(Y/X)** este perspectiva matematică condițională a valorii funcției corespunzător valorii particulare **X = x**.

Linia medie empirică, numită **linia empirică de regresie**, se va determina cu expresia:

$$Y = \bar{y} + r \frac{s_y}{s_x}(x - \bar{x}) \tag{7.31}$$

sau:

$$Y = a + b.x \tag{7.32}$$

unde:

$$b = r \frac{s_y}{s_x} \tag{7.33}$$

respectiv:

$$a = \bar{y} - b.\bar{x} \tag{7.34}$$

Parametrul **b** se numeşte **coeficient de regresie**.

Pentru majoritatea problemelor practice se poate accepta că dispersia condițională a valorii **Y** nu depinde de variația lui **x**, adică ecuația (7.29) va deveni:

$$\sigma_y^2 = f_2(x) = const. \tag{7.35}$$

în care caz parametrii $\bar{x}$ şi $\bar{y}$ din egalitatea (7.31) pot fi determinați cu expresiile (7.20)...(7.23). Pentru evaluarea dispersiei (7.35) se foloseşte dispersia empirică $s_{y/x}^2$ care se determină cu o expresie similară dispersiei condiționale (7.25):

$$s_{y/x}^2 = \frac{\sum_{i=1}^{m}(n_{i\bullet} - 1)s_{y/x_i}^2}{n - m} \tag{7.36}$$

Material cu drept de autor

unde: $n = \sum_{i=1}^{m} n_{i\bullet}$ - numărul total al experiențelor.

Acceptarea condiției (7.35) şi ca urmare, utilizării expresiei (7.36), trebuie să preceadă verificarea omogenității dispersiei condiționale (7.35).

Abaterea medie pătratică $s_{y/x} = \sqrt{s^2_{y/x}}$ se foloseşte în calitate de măsură a împrăştierii rezultatelor observațiilor în jurul liniei de regresie.

În cazul general, la analiza regresiei între două variabile **X** şi **Y**, se definesc două linii de regresie, regresia **Y** în funcție de **X** şi regresia **X** în funcție de **Y**. În primul caz, s-a definit prin expresia (7.28) media $m_y = f_1(x)$, vom defini acum şi media:

$$m_x = f(y) \tag{7.37}$$

deşi, în practică, prezintă importanță numai una dintre ele.

Evaluarea liniei empirice de regresie **X** în raport cu **Y**, se scrie analog expresiei (7.31)

$$X = \bar{x} + r\frac{s_x}{s_y}(y - \bar{y}) \tag{7.38}$$

sau:

$$X = a_1 + b_1.y \tag{7.39}$$

unde:

$$b_1 = r\frac{s_x}{s_y} \tag{7.40}$$

respectiv:

$$a_1 = \bar{x} - b_1.\bar{y} \tag{7.41}$$

Evaluarea dispersiei condiționale se va determina cu expresia:

$$s^2_{x/y} = \frac{\sum_{j=1}^{k}(n_{\bullet j} - 1)s^2_{x/y_j}}{n - k} \tag{7.42}$$

Şi în acest caz, în prealabil se va analiza omogenitatea dispersiilor cu criteriul lui Barttlet.

În general, linia de regresie a lui **Y** în dependență de **X** nu coincide cu linia de regresie **X** în funcție de **Y**. Ele se vor intersecta în punctul cu coordonatele $\bar{x}$ şi $\bar{y}$, iar unghiul dintre ele este cu atât mai mare, cu cât mai

Material cu drept de autor

mult coeficientul de corelație diferă de unitate. Când $|r|$ = 1, liniile de regresie vor coincide.

În general, metodologia practică de determinare a poziției liniei de regresie se face prin metoda celor mai mici pătrate.

În acest scop, se consideră reprezentarea grafică din figura 7.2.

Se scrie ecuația dreptei de regresie cu expresia (7.32): $Y = a + b.x$ care este estimata empirică a ecuației teoretice de regresie:

$$f(x,\alpha,\beta) = \alpha + \beta \cdot x$$

Ca urmare, în ecuația (6.32), variabila **Y** este funcția de regresie empirică care estimează pe $f(x,\alpha,\beta) = \alpha + \beta \cdot x$. Valorile **x** sunt considerate exacte, nesupuse la erori. În practică, erorile lui **x** sunt foarte mici comparativ cu fluctuațiile valorilor **y**.

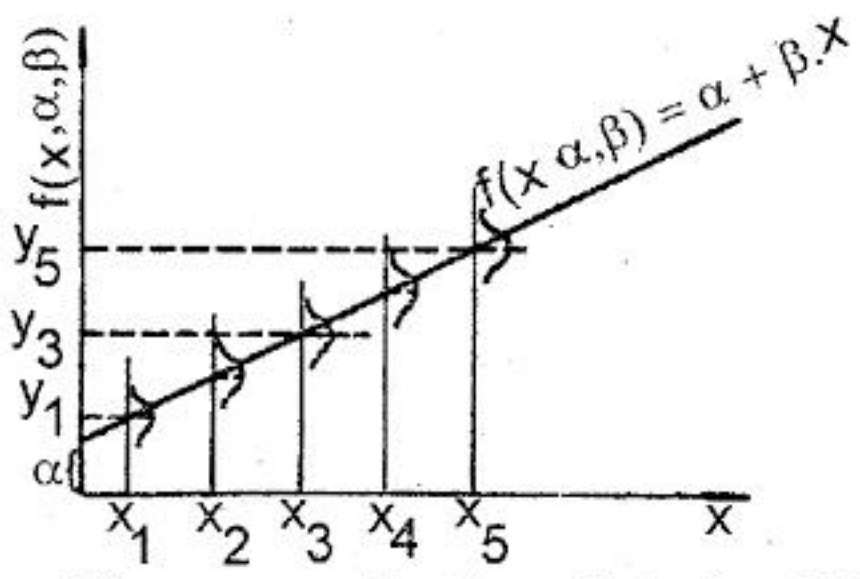


Fig.7.2. Repartiția valorilor y_i în funcție de x_i

Pentru ca funcția **Y** să fie dreapta căutată este necesar să se îndeplinească condiția: **suma pătratelor abaterilor față de ea, a tuturor valorilor de observație y, să fie minimă**. Se va scrie suma:

$$S = \Sigma(y_i - Y_{x_i})^2 = \Sigma(y_i - a - b.x_i)^2 \to \min \qquad (7.43)$$

Se anulează primele derivate parțiale în raport cu **a** şi **b**. Va rezulta sistemul de ecuații (7.45):

$$\Rightarrow \begin{cases} \dfrac{\partial S}{\partial a} = 0 \\ \dfrac{\partial S}{\partial b} = 0 \end{cases} \Rightarrow \begin{cases} 2\Sigma(y_i - a - b.x_i)(-1) = 0 \\ 2\Sigma(y_i - a - b.x_i)(-x_i) = 0 \end{cases} \Rightarrow \qquad (7.44)$$

$$\Rightarrow \begin{cases} n.a + b\Sigma x_i = \Sigma y_i \\ a\Sigma x_i + b\Sigma x_i^2 = \Sigma x_i y_i \end{cases} \qquad (7.45)$$

Observație. Relațiile sunt valabile şi în cazul când pentru aceeaşi valoare **x** avem mai multe valori **y**. În acest caz - ca şi în cel tratat - prin **n** se

Material cu drept de autor

înțelege numărul total al observațiilor, iar operatorii Σ include, de asemenea, toate valorile **x** şi respectiv **y**.

În ecuațiile obişnuite, numite normale, necunoscutele sunt **a** şi **b**. Sumele în care intră **x** şi **y** sunt cunoscute.

Se elimină necunoscuta **a** între ecuațiile (7.45) şi se obține:

$$b = \frac{n\sum x.y - \left(\sum x\right)\left(\sum y\right)}{n\sum x^2 - \left(\sum x\right)^2} \qquad (7.46)$$

sau:

$$b = \frac{\Sigma xy - \dfrac{(\Sigma x)(\Sigma y)}{n}}{\Sigma x^2 - \dfrac{(\Sigma x)^2}{n}} \qquad (7.47$$

Valoarea **a** se obține prin înlocuirea lui **b** astfel calculat în prima ecuație normală (7.45):

$$a = \frac{\Sigma y - b\Sigma x}{n} \qquad (7.48)$$

sau:

$$a = \frac{(\Sigma x)(\Sigma y) - (\Sigma y)\left(\Sigma x^2\right)}{(\Sigma x)^2 - n\left(\Sigma x^2\right)} \qquad (7.49)$$

Fie acum $\bar{x}$ şi $\bar{y}$ mediile valorilor **x** şi respectiv **y**. Se observă uşor că $\Sigma x = \mathbf{n.\bar{x}}$ şi $\Sigma y = \boldsymbol{n.\bar{y}}$. Ecuațiile (7.46) şi (7.48) devin:

$$b = \frac{\Sigma x.y - n.\bar{x}.\bar{y}}{\Sigma x^2 - n.x^2} \qquad (7.50)$$

$$a = \bar{y} - b.\bar{x} \qquad (7.51)$$

Relațiile (7.50) şi (7.51) ne asigură că punctul de coordonare $\bar{x}$ şi $\bar{y}$ se găseşte pe dreapta definită de ecuația: $\boldsymbol{Y = a + b.x}$

Dacă în ecuația dreptei de regresie **Y = a + b.x**, se înlocuieşte **a** cu $\bar{y}$ - **b.** $\bar{x}$ din relația (7.51), vom avea:

$$Y = \bar{y} - b.\bar{x} + b.x \qquad (7.52)$$

sau:

$$Y = \bar{y} + b.\left(x - \bar{x}\right) \qquad (7.53)$$

Material cu drept de autor

unde dreapta de regresie se exprimă în raport cu **b** şi cu valorile medii generale $\overline{y}$ şi $\overline{x}$, formă utilă în unele calcule.

Dacă înlocuim în relația (6.53) pe **x** cu (**x**-$\overline{x}$) şi pe **y** cu (**y**-$\overline{y}$), adică dacă deplasăm originea axelor de coordonate în punctul de coordonate **x** = $\overline{x}$ şi **y** = $\overline{y}$, valoarea **b** nu se modifică deoarece panta dreptei de regresie rămâne aceeaşi. Rezultă:

$$b = \frac{\Sigma(x-\overline{x})(y-\overline{y})}{\Sigma(x-\overline{x})^2} \tag{7.54}$$

Alte forme pentru **b** şi **a**, în cazul când numai originea axei **ox** s-a deplasat în $\overline{x}$ sunt:

$$b = \frac{\Sigma(x-\overline{x})\overline{y}}{\Sigma(x-\overline{x})^2} \tag{7.55}$$

$$a = \frac{\Sigma y}{n} = \overline{y} \tag{7.56}$$

Se ajunge la aceste relații înlocuind în ecuațiile (7.45) valoarea **x** cu (**x**-$\overline{x}$). Rezultă:

$$n.a + b\Sigma(x-\overline{x}) = \Sigma y \tag{7.57}$$

$$a\Sigma(x-\overline{x}) + b\Sigma(x-\overline{x})^2 = \Sigma(x-\overline{x}).y \tag{7.58}$$

Se rezolvă apoi sistemul format, în raport cu **a** sau **b**, după aceeaşi metodologie.

Valoarea coeficienților **a** şi **b** se pot calcula şi prin alte metode. De exemplu, cu ajutorul determinanților se obține:

$$a = \frac{\Delta_a}{\Delta} = \frac{\begin{vmatrix} \Sigma y_i & \Sigma x_i \\ \Sigma x_i y_i & \Sigma x_i^2 \end{vmatrix}}{\begin{vmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{vmatrix}} = \frac{\Sigma x_i^2 \Sigma y_i - \Sigma x_i \Sigma x_i y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} \tag{7.59}$$

Material cu drept de autor

$$b = \frac{\Delta_b}{\Delta} = \frac{\begin{vmatrix} n & \Sigma y_i \\ \Sigma x_i & \Sigma x_i y_i \end{vmatrix}}{\begin{vmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{vmatrix}} = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} \qquad (7.60)$$

Cazul ($\mathbf{x_i}$, $\mathbf{y_i}$, $\mathbf{n_i}$) – ponderat cu „$\mathbf{n_i}$”:

$$a = \frac{\Sigma x_i^2 n_i \Sigma y_i n_i - \Sigma x_i n_i \Sigma x_i y_i n_i}{\Sigma n_i \Sigma x_i^2 n_i - (\Sigma x_i n_i)^2} \qquad (7.61)$$

$$b = \frac{\Sigma n_i \Sigma x_i y_i n_i - \Sigma x_i n_i \Sigma y_i n_i}{\Sigma n_i \Sigma x_i^2 n_i - (\Sigma x_i n_i)^2} \qquad (7.62)$$

Cazul tabelului de corelație:

$$a = \frac{\Sigma x_i^2 n_{i\bullet} \Sigma y_j n_{\bullet j} - \Sigma x_i n_{i\bullet} \Sigma x_i y_j n_{ij}}{\Sigma\Sigma n_{ij} \Sigma x_i^2 n_{i\bullet} - (\Sigma x_i n_{i\bullet})^2} \qquad (7.63)$$

$$b = \frac{\Sigma\Sigma n_{ij} \Sigma\Sigma x_i y_j n_{ij} - \Sigma x_i n_{i\bullet} \Sigma y_j n_{\bullet j}}{\Sigma\Sigma n_{ij} \Sigma x_i^2 n_{i\bullet} - (\Sigma x_i n_{i\bullet})^2} \qquad (7.64)$$

La trecerea de la cazul ($\mathbf{x_i}$, $\mathbf{y_i}$, $\mathbf{n_i}$) la cazul tabelului de corelație s-au utilizat substituțiile:

$$\Sigma n_i \to \Sigma\Sigma n_{ij};\ \Sigma x_i^k n_i \to \Sigma x_i^k n_{i\bullet}$$

$$\Sigma y_i^k n_i \to \Sigma y_j^k n_{\bullet j};\ \Sigma x_i y_i n_i \to \Sigma\Sigma x_i y_j n_{ij}$$

Pentru rezolvarea ecuațiilor normale, foarte comodă s-a dovedit a fi aranjarea datelor sub formă de tabel. În cazul în care datele inițiale conțin cifre mari este avantajos să se lucreze cu cifre micşorate cu o constantă care nu trebuie să fie însă aceeaşi pentru variabilele **x** şi **y**. Drept constantă cu care se micşorează datele inițiale este preferată media lor aritmetică ceea ce înseamnă că calculele se vor realiza cu abaterea observațiilor individuale de la media lor aritmetică, aducă cu valorile $(x_i - \bar{x})$ şi respectiv cu $(y_i - \bar{y})$.

Exemplul de calcul 7.2. Pe un lot de teren agricol experimental se cercetează dependența dintre creşterea cantității relative de substanță activă

Material cu drept de autor

utilizată ca îngrăşământ şi creşterea cantității relative a producției agricole. Rezultatele obținute se prezintă în tabelul 7.3.

Tabelul 7.3

Date experimentale de calcul:

Creşterea relativă de substanță activă **x**	Creşterea relativă a producției agricole **y**	x^2	y^2	$x.y$
0,20	0,15	0,04	0,0225	0,030
0,40	0,30	0,16	0,0900	0,120
0,60	0,44	0,36	0,1936	0,264
0,80	0,61	0,64	0,3721	0,488
1,00	0,74	1,00	0,5476	0,740
$\Sigma x = 3{,}00$ $\bar{x} = 0{,}60$ $(\Sigma x)^2 = 9$	$\Sigma y = 2{,}24$ $\bar{y} = 0{,}448$	$\Sigma x^2 = 2{,}20$	$\Sigma y^2 = 1{,}2258$	$\Sigma x.y = 1{,}642$

Se cere să se stabilească dreapta de regresie în acest caz.

Rezolvare. Folosind relațiile de calcul (7.46) şi (7.48) se obține:

$$b = \frac{n\Sigma x.y - (\Sigma x)(\Sigma y)}{n\Sigma x^2 - (\Sigma x)^2} = \frac{5.1{,}642 - 3.2{,}24}{5.2{,}20 - 9} = 0{,}745$$

$$a = \frac{\Sigma y - b\Sigma x}{n} = \frac{2{,}24 - 0{,}745.3}{5} = 0{,}001$$

şi deci: $Y = a + b.x = 0{,}001 + 0{,}745.x$

Estimarea dispersiilor: întregii populații a valorilor **y,** parametrilor **a** şi **b,** dreptei de regresie **Y** se realizează prin metodologii adecvate.

Între coeficienții de regresie şi coeficienții de corelație se verifică următoarele legături de dependență.

În cazul regresiei lui **y** în funcție de **x**, între coeficientul de regresie $\mathbf{b_{y/x}}$ şi coeficientul de corelație $\mathbf{r_{y/x}}$ există următoarea relație:

$$b_{y/x} = r_{y/x}\frac{\sigma_y}{\sigma_x} \tag{7.65}$$

de unde:

$$r_{y/x} = b_{y/x}\frac{\sigma_x}{\sigma_y} \tag{7.66}$$

În mod analog, în cazul regresiei lui **x** în funcție de **y** se verifică relațiile:

Material cu drept de autor

$$b_{x/y} = r_{x/y} \frac{\sigma_x}{\sigma_y} \tag{7.67}$$

de unde:

$$r_{x/y} = b_{x/y} \frac{\sigma_y}{\sigma_x} \tag{7.68}$$

Din relațiile de mai sus rezultă că un coeficient de corelație poate fi egal cu coeficientul de regresie corespunzător atunci când dispersiile celor două variabile corelate sunt egale:

$$r_{y/x} = b_{y/x} \text{ când } \sigma_x = \sigma_y \text{ şi } r_{x/y} = b_{x/y} = b_{y/x} \tag{7.69}$$

Între coeficienții de regresie $\mathbf{b_{y/x}}$, $\mathbf{b_{x/y}}$ şi coeficientul de corelație **r** există următoarea relație:

$$r_{x/y} = \sqrt{b_{y/x} \times b_{x/y}} \tag{7.70}$$

Această ultimă expresie permite evaluarea unuia dintre cele trei elemente pe baza celorlalte două. Astfel, în mod obişnuit se cunoaşte $\mathbf{b_{x/y}}$ şi **r** şi pe baza acestora se determină $\mathbf{b_{y/x}}$.

7.7. Corelație şi regresie liniară multiplă

În practica cercetărilor fenomenelor economico-sociale se întâlnesc cazuri când caracteristica care interesează depinde esențial nu numai de una, ci de două sau mai multe caracteristici. Corelația acestor caracteristici se numeşte multiplă. În funcție de numărul caracteristicilor există legătură bidimensională, tridimensională, cuadrimensională etc.

Corelația multiplă, de orice fel (dublă, triplă etc.), poate fi în unele cazuri liniară, în altele neliniară (curbe de gradul doi, trei şi mai mare).

Se va considera legătura liniară dublă, dintre o variabilă factorială **y** şi două variabile independente $\mathbf{x_1}$ şi $\mathbf{x_2}$. Verificarea prezenței acestei legături, precum şi măsurarea intensității sale, se pot face prin evaluarea coeficientului de corelație multiplă liniară care se determină cu ajutorul următoarei relații de calcul:

$$r_{y_{x_1x_2}} = \sqrt{\frac{r_{y_{x_1}}^2 + r_{y_{x_2}}^2 - 2 \cdot r_{y_{x_1}} r_{y_{x_2}} r_{x_1x_2}}{1 - r_{x_1x_2}^2}} \tag{7.71}$$

Material cu drept de autor

în care **y** este parametrul urmărit, dependent de $\mathbf{x_1}$ şi $\mathbf{x_2}$, $r_{y_{x_i}}$ este evaluarea coeficientului de corelație pentru $\mathbf{x_1}$ şi $\mathbf{x_2}$, iar $r_{x_1x_2}$ este evaluarea coeficientului de corelație pentru $\mathbf{x_1}$ şi $\mathbf{x_2}$. În care:

$$r_{y_{x_i}} = b_{y_{x_i}} \frac{\sigma_{x_i}}{\sigma_y} \tag{7.72}$$

unde: $r_{y_{x_i}}$ - coeficientul de corelație a lui **y** în funcție de $\mathbf{x_i}$; $b_{y_{x_i}}$ - coeficientul de regresie.

Coeficientul multiplu de corelație **r** este întotdeauna pozitiv. Valoarea sa numerică variază de la **0** la **1**. Dacă legătura liniară a lui **y** cu $\mathbf{x_1}$ şi $\mathbf{x_2}$ lipseşte, speranța matematică a valorii **r** este egală cu zero $[M(r)=0]$; poate să existe însă o legătură neliniară. În cazul legăturii liniare exacte (funcționale) **y** cu $\mathbf{x_1}$ şi $\mathbf{x_2}$, **M(r) = 1**. În toate celelalte cazuri $\mathbf{0 < r < 1}$.

Corelație parțială. Dacă se are în vedere că în cazul corelației simple se studiază legătura liniară dintre doi factori, neglijându-se influența celorlalți factori care acționează concomitent asupra variabilei **y**, în cazul corelației multiple liniare se studiază influența simultană a două sau mai multe caracteristici factoriale asupra caracteristicii rezultative, atunci ***influența exercitată de o variabilă independentă asupra variabilei dependente, presupunând că celelalte variabile independente se mențin la nivel constant, este studiată de corelația parțială***. Intensitatea corelației parțiale se apreciază cu ajutorul coeficientului de corelație parțială de ordinul **l** care se determină, pentru cazul general, cu expresia:

$$r_{yx_1/x_2,x_3,\ldots,x_k} = \frac{r_{y/x_1,x_2,\ldots,x_{k-1}} - r_{yx_k/x_2,x_3,\ldots,x_{k-1}} \times r_{x_1x_k/x_2,x_3,\ldots,x_{k-1}}}{\sqrt{(1-r^2_{yx_k/x_2,x_3,\ldots,x_{k-1}})(r^2_{x_1x_k/x_2,x_3,\ldots,x_{k-1}})}} \tag{7.73}$$

unde: $r_{yx_1/x_2,x_3,\ldots,x_k}$ reprezintă coeficientul de corelație parțială a lui **y** şi $\mathbf{x_1}$ când $\mathbf{x_2, x_3,\ldots,x_k}$ sunt constante; $r_{yx_k/x_2,x_3,\ldots,x_{k-1}}$ - reprezintă coeficientul de corelație parțială a lui **y** şi $\mathbf{x_k}$ când $\mathbf{x_2, x_3,\ldots,x_{k-1}}$ sunt constante şi $r_{x_1x_k/x_2,x_3,\ldots,x_{k-1}}$ - reprezintă coeficientul de corelație parțială a lui $\mathbf{x_1}$ şi $\mathbf{x_2}$ când $\mathbf{x_2, x_3,\ldots,x_{k-1}}$ sunt constante.

În cazul a două variabile independente $\mathbf{x_1}$ şi $\mathbf{x_2}$ şi a unei variabile dependente **y**, coeficientul de corelație parțială de ordinul **l** se calculează cu expresiile:

a. coeficientul de corelație parțială dintre **y** şi $\mathbf{x_1}$, neglijându-se influența lui $\mathbf{x_2}$:

Material cu drept de autor

$$r_{yx_1/x_2} = \frac{r_{yx_1} - r_{yx_2}.r_{x_1x_2}}{\sqrt{(1-r_{yx_2}^2).(1-r_{x_1x_2}^2)}} \quad (7.74)$$

b. coeficientul de corelație parțială dintre **y** şi $\mathbf{x_2}$, neglijându-se influența lui $\mathbf{x_1}$:

$$r_{yx_2/x_1} = \frac{r_{yx_2} - r_{yx_1}.r_{x_1x_2}}{\sqrt{(1-r_{yx_1}^2).(1-r_{x_1x_2}^2)}} \quad (7.75)$$

În cazul a trei variabile independente $\mathbf{x_1}$, $\mathbf{x_2}$, $\mathbf{x_3}$ şi a unei variabile dependente **y**, coeficienții de corelație parțială de ordinul al **II**-lea se pot determina cu ajutorul coeficienților de corelație parțială de ordinul **I**, cu expresiile:

a. coeficientul de corelație parțială dintre **y** şi $\mathbf{x_1}$, neglijându-se influența lui $\mathbf{x_2}$ şi $\mathbf{x_3}$:

$$r_{yx_1/x_2x_3} = \frac{r_{yx_1/x_2} - r_{yx_3/x_2}.r_{x_1x_3/x_2}}{\sqrt{(1-r_{yx_3/x_2}^2).(1-r_{x_1x_3/x_2}^2)}} \quad (7.76)$$

b. coeficientul de corelație parțială dintre **y** şi $\mathbf{x_2}$, neglijându-se influența lui $\mathbf{x_1}$ şi $\mathbf{x_3}$:

$$r_{yx_2/x_1x_3} = \frac{r_{yx_2/x_1} - r_{yx_3/x_1}.r_{x_2x_3/x_1}}{\sqrt{(1-r_{yx_3/x_1}^2).(1-r_{x_2x_3/x_1}^2)}} \quad (7.77)$$

c. coeficientul de corelație parțială dintre **y** şi $\mathbf{x_3}$, neglijându-se influența lui $\mathbf{x_2}$ şi $\mathbf{x_1}$:

$$r_{yx_3/x_1x_2} = \frac{r_{yx_3/x_1} - r_{yx_2/x_1}.r_{x_2x_3/x_1}}{\sqrt{(1-r_{yx_2/x_1}^2).(1-r_{x_2x_3/x_1}^2)}} \quad (7.78)$$

Regresia multiplă. Regresia multiplă de orice fel (dublă, triplă etc.), poate fi în unele cazuri liniară, în altele neliniară (curbe de gradul doi, trei şi mai mare).

Considerăm regresia liniară multiplă:

$$Y_{x_1,x_2,\ldots,x_k} = a_0 + a_1x_1 + a_2x_2 + \ldots + a_kx_k \quad (7.79)$$

Parametrii $\mathbf{a_0}$, $\mathbf{a_1}$, $\mathbf{a_2,\ldots,a_k}$ se determină prin metoda celor mai mici pătrate:

$$\sum_{i=1}^{k}(y_i - Y_{x_1,x_2,\ldots,x_k})^2 = \min \quad (7.80)$$

care conduce la sistemul de ecuații normale:

Material cu drept de autor

$$\begin{cases} n.a_0 + a_1\Sigma x_{1i} + \ldots + a_k\Sigma x_{ki} = \Sigma y_i \\ a_0\Sigma x_{1i} + a_1\Sigma x_{1i}^2 + \ldots + a_k\Sigma x_{1i}x_{ki} = \Sigma x_{1i}y_i \\ a_0\Sigma x_{2i} + a_1\Sigma x_{1i}x_{2i} + \ldots + a_k\Sigma x_{2i}x_{ki} = \Sigma x_{2i}y_i \\ \ldots\ldots\ldots\ldots\ldots \\ a_0\Sigma x_{ki} + a_1\Sigma x_{1i}x_{ki} + \ldots + a_k\Sigma x_{ki}^2 = \Sigma x_{ki}y_i \end{cases} \quad (7.81)$$

Exemplu de calcul 7.3. Se dau datele din tabelul următor:

Tabelul 7.4

y	14	43,5	18,5	30	40	51,5	73
x_1	4	5	7	10	12	15	20
x_2	2	10	2	4	6	8	12

Se cere:

a. Determinarea prin metoda celor mai mici pătrate a ecuației de regresie: $y = a_0 + a_1x_1 + a_2x_2$;

b. Determinarea valorilor estimate pentru **y, x_1** şi **x_2**;

c. Determinarea coeficienților de corelație: r_{y/x_1}, r_{y/x_2}, r_{x_1/x_2}.

d. Determinarea coeficienților de corelație multiplă liniară **y, x_1** şi **x_2**;

e. Determinarea coeficienților de corelație parțială liniară r_{yx_1/x_2}, r_{yx_2/x_1}.

Rezolvare:

a. Se urmăreşte metodologia de calcul:

$$y_{x_i} = a_0 + a_1.x_1 + a_2.x_2$$

$$S = \Sigma(y - y_{x_i}) = \Sigma(y - a_0 - a_1.x_1 - a_2.x_2)^2 \Rightarrow \min$$

$$\begin{cases} \dfrac{\partial S}{\partial a_0} = 0 \\ \dfrac{\partial S}{\partial a_1} = 0 \\ \dfrac{\partial S}{\partial a_2} = 0 \end{cases} \Rightarrow \begin{cases} 2\Sigma(y - a_0 - a_1x_1 - a_2x_2)(-1) = 0 \\ 2\Sigma(y - a_0 - a_1x_1 - a_2x_2)(-x_1) = 0 \\ 2\Sigma(y - a_0 - a_1x_1 - a_2x_2)(-x_2) = 0 \end{cases} \Rightarrow$$

Material cu drept de autor

$$\Rightarrow \begin{cases} \Sigma y_i = a_0 n - a_1 \Sigma x_{1_i} - a_2 \Sigma x_{2_i}) \\ \Sigma y_i . x_{1_i} = a_0 \Sigma x_{1_i} - a_1 \Sigma x_{1_i}^2 - a_2 \Sigma x_{1_i} x_{2_i}) \\ \Sigma y_i . x_{2_i} = a_0 \Sigma x_{2_i} - a_1 \Sigma x_{1_i} x_{2_i}) - a_2 \Sigma x_{2_i}^2 \end{cases}$$

Se întocmeşte tabelul 7.5 de calcul:

Tabelul 7.5

y	x_1	x_2	$y.x_1$	$y.x_2$	$x_1.x_2$	x_1^2	x_2^2
14	4	2	56	28	8	16	4
43,5	5	10	217,5	435	50	25	100
18,5	7	2	129,5	37	14	49	4
30	10	4	300	120	40	100	16
40	12	6	480	240	72	144	36
51,5	15	8	772,5	412	120	225	64
73	20	12	1460	876	240	400	144
Σ :**270,5**	Σ :**73**	Σ :**44**	Σ :**3415,5**	Σ :**2148**	Σ :**544**	Σ :**959**	Σ :**368**

Utilizând datele din tabelul 6.5, sistemul de ecuații normale devine:

$$\begin{cases} 270,5 = 7a_0 - 73a_1 - 44a_2 \\ 3415,5 = 73a_0 - 959a_1 - 544a_2 \\ 2148 = 44a_0 - 544a_1 - 368a_2 \end{cases}$$

După rezolvare rezultă: $a_0 = 0,98 \;\; a_1 = 1,5 \;\; a_2 = 3,50$

Deci: $y_{x_i} = 0,98 + 1,50.x_1 + 3,50.x_2$

b. Determinarea valorilor estimate pentru **y, x_1** şi **x_2**. Se realizează tabelul 7.6.

c. Determinarea coeficienților de regresie: r_{y/x_1}, r_{y/x_2} , r_{x_1/x_2}.

c_1. Se scrie ecuația:

$$r_{y/x_1} = \frac{n\Sigma y.x_1 - (\Sigma y).(\Sigma x_1)}{\sqrt{[n\Sigma y^2 - (\Sigma y)^2].[n\Sigma x_1^2 - (\Sigma x_1)^2]}}$$

Introducând datele din tabelul 6.7 rezultă:

Material cu drept de autor

$$r_{y/x_1} = \frac{7 \times 3415{,}5 - 270{,}5 \times 73}{\sqrt{[7 \times 12911{,}75 - 270{,}5^2].[7 \times 959 - 73^2]}} = 0{,}8527$$

Tabelul 7.6

Calculul sumelor relevante

y	14	43,5	18,5	30	40	51,5	73
x_1	4	5	7	10	12	15	20
x_2	2	10	2	4	6	8	12
$y_{x_i\ estimat}$	14,016	43,585	18,531	30,062	40,088	51,619	73,176

c₂.

$$r_{y/x_2} = \frac{n\Sigma y.x_2 - (\Sigma y).(\Sigma x_2)}{\sqrt{[n\Sigma y^2 - (\Sigma y)^2].[n\Sigma x_2^2 - (\Sigma x_2)^2]}}$$

$$r_{y/x_2} = \frac{7 \times 2148 - 270{,}5 \times 44}{\sqrt{17212 \times 640}} = 0{,}9442$$

c₃.

$$r_{x_1/x_2} = \frac{n\Sigma x_1.x_2 - (\Sigma x_1).(\Sigma x_2)}{\sqrt{[n\Sigma x_1^2 - (\Sigma x_1)^2].[n\Sigma x_2^2 - (\Sigma x_2)^2]}}$$

$$r_{x_1/x_2} = \frac{7 \times 544 - 73 \times 44}{\sqrt{1384 \times 640}} = 0{,}6332$$

d. Determinarea coeficienților de corelație multiplă liniară **y, x_1 şi x_2**;

d₁.

$$r_{y/x_1x_2} = \sqrt{\frac{r_{y/x_1}^2 + r_{y/x_2}^2 - 2.r_{y/x_1}.r_{y/x_2}.r_{x_1/x_2}}{1 - r_{x_1/x_2}^2}}$$

$$r_{y/x_1x_2} = \sqrt{\frac{0{,}8527^2 + 0{,}9442^2 - 2.0{,}8527.0{,}9442.0{,}6332}{1 - 0{,}6332^2}} = 0{,}9999582$$

d₂.

$$r_{x_1/yx_2} = \sqrt{\frac{r_{y/x_1}^2 + r_{x_1/x_2}^2 - 2.r_{y/x_1}.r_{y/x_2}.r_{x_1/x_2}}{1 - r_{y/x_2}^2}}$$

Material cu drept de autor

$$r_{x_1/yx_2} = \sqrt{\frac{0{,}8527^2 + 0{,}6332^2 - 2.0{,}8527.0{,}9442.0{,}6332}{1 - 0{,}9442^2}} = 0{,}9997695$$

d3.

$$r_{x_2/yx_1} = \sqrt{\frac{r_{y/x_2}^2 + r_{x_1/x_2}^2 - 2.r_{y/x_1}.r_{y/x_2}.r_{x_1/x_2}}{1 - r_{y/x_1}^2}}$$

$$r_{x_2/yx_1} = \sqrt{\frac{0{,}9442^2 + 0{,}6332^2 - 2.0{,}8527.0{,}9442.0{,}6332}{1 - 0{,}8527^2}} = 0{,}9999083$$

e. Determinarea coeficienților de corelație parțială liniară

e1.

$$r_{yx_1/x_2} = \frac{r_{yx_1} - r_{yx_2}.r_{x_1x_2}}{\sqrt{(1 - r_{yx_2}^2).(1 - r_{x_1x_2}^2)}}$$

$$r_{yx_1/x_2} = \frac{0{,}8527 - 0{,}9442 \times 0{,}6332}{\sqrt{(1 - 0{,}9442^2).(1 - 0{,}6332^2)}} = 0{,}9996159$$

e2.

$$r_{yx_2/x_1} = \frac{r_{yx_2} - r_{yx_1}.r_{x_1x_2}}{\sqrt{(1 - r_{yx_1}^2).(1 - r_{x_1x_2}^2)}}$$

$$r_{yx_2/x_1} = \frac{0{,}9442 - 0{,}8527 \times 0{,}6332}{\sqrt{(1 - 0{,}8527^2).(1 - 0{,}6332^2)}} = 0{,}9998474$$

7.8. Raportul de corelație şi coeficientul de determinație

Raportul de corelație este un indicator ca măsoară intensitatea legăturilor dintre variabilele statistice, indiferent de forma acestor legături – liniare sau neliniare – pe baza descompunerii variației totale a lui **y** în cele două componente, cea sistematică şi cea aleatoare.

Se determină cu expresia:

Material cu drept de autor

$$R = \sqrt{1 - \frac{\Sigma(y - \hat{y})^2}{\Sigma(y - \bar{y})^2}} \qquad 0 \le R \le 1 \tag{7.82}$$

unde numărătorul fracției măsoară variația datorată factorilor aleatori, iar numitorul se referă la variația totală.

Dacă funcția de regresie este lianiară, raportul de corelație se mai poate calcula şi cu expresia:

$$R = \sqrt{1 - \frac{\Sigma y^2 - a\Sigma y - b\Sigma x.y}{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}} \tag{7.83}$$

sau, pentru regresia liniară multiplă:

$$R = \sqrt{\frac{a_0\Sigma y + a_1\Sigma y x_1 + \ldots + a_k\Sigma y x_k - \frac{(\Sigma y)^2}{n}}{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}} \tag{7.84}$$

Prin transformări succesive se poate ajunge la diferite forme ale ecuației raportului de corelație:

$$R = \sqrt{1 - \frac{\Sigma(y - y_x)^2}{\Sigma(y_i - \bar{y})^2}} \quad \text{sau} \quad R = \sqrt{\frac{a\Sigma y_i + b\Sigma x_i y_i - \frac{(\Sigma y_i)^2}{n}}{\Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n}}} \tag{7.85}$$

Dacă perechile de valori corelate sunt ponderate cu $\mathbf{n_i}$ atunci formula raportului de corelație devine:

$$R = \sqrt{\frac{a\Sigma y_i n_i + b\Sigma x_i y_i n_i - \frac{(\Sigma y_i n_i)^2}{\Sigma n_i}}{\Sigma y_i^2 n_i - \frac{(\Sigma y_i n_i)^2}{\Sigma n_i}}} \tag{7.86}$$

În cazul unor distribuții bidimensionale ($\mathbf{x_i}$, $\mathbf{y_i}$, $\mathbf{n_{ij}}$) se aplică o formulă modificată care să țină seama de indicii de variație ai celor două variabile şi de frecvența perechilor de valori corelate $\mathbf{n_{ij}}$:

$$R = \sqrt{\frac{a\Sigma y_j n_{\bullet j} + b\Sigma\Sigma x_i y_j n_{ij} - \frac{(\Sigma y_j n_{\bullet j})^2}{\Sigma\Sigma n_{ij}}}{\Sigma y_j^2 n_{\bullet j} - \frac{(\Sigma y_j n_{\bullet j})^2}{\Sigma\Sigma n_{ij}}}} \tag{7.87}$$

Material cu drept de autor

Coeficientul de determinație (R^2) este pătratul coeficientului de corelație simplă sau multiplă şi exprimă ponderea cu care influențează caracteristica sau caracteristicile factoriale incluse în model, asupra caracteristicii rezultative.

Determinația multiplă se poate calcula şi pe baza raportului de corelație multiplă liniară. Astfel, pentru două variabile independente:

$$R^2 = \frac{a_1\left(n\Sigma yx_1 - \Sigma x_1\Sigma y\right)}{n\Sigma y^2 - \left(\Sigma y\right)^2} + \frac{a_2\left(n\Sigma yx_2 - \Sigma x_2\Sigma y\right)}{n\Sigma y^2 - \left(\Sigma y\right)^2} \qquad (7.88)$$

unde cei doi termeni reprezintă coeficienții de determinație parțială, reflectând influența fiecăreia dintre cele două variabile independente. În general, pentru **k** variabile, coeficienții de determinație se exprimă prin:

$$R^2 = \frac{a_0\Sigma y + a_1\Sigma yx_1 + a_2\Sigma yx_2 + \ldots + a_k\Sigma yx_k - \dfrac{\left(\Sigma y\right)^2}{n}}{n\Sigma y^2 - \left(\Sigma y\right)^2} \qquad (7.89)$$

$$R^2 = \frac{a\left(n\Sigma yx - \Sigma x\Sigma y\right)}{n\Sigma y^2 - \left(\Sigma y\right)^2} \qquad (7.90)$$

Suma coeficienților de determinație parțială este egală cu coeficientul de determinație totală.

Exemplu de calcul 7.4. Ştiind că între cantitatea de substanță activă utilizată ca îngrăşământ şi cantitatea de producție (codificate) s-au înregistrat valorile din coloanele 1 şi 2 din tabelul 7.7, apreciate ca dependență liniară, se cere să se determine indicatorii de corelație şi regresie.

Tabelul 7.7

x	y	x.y	x^2	$\hat{y}$	$(y-\hat{y})^2$	$(y-\bar{y})^2$	y^2
1	2	3	4	5	6	7	8
0	2	-	-	2,10	0,0100	10,24	4
1	3,5	3,5	1	3,65	0,0225	2,89	12,25
2	5,5	11	4	5,20	0,0900	0,04	30,25
3	7	21	9	6,75	0,0625	3,24	49
4	8	32	16	8,30	0,0900	7,84	64
10	26	67,5	30	26	0,275	24,25	159,5

Rezolvare. Ecuația dreptei de regresie liniară care descrie legătura dintre două variabile (**y** şi **x**), considerând că ceilalți factori au o acțiune constantă şi neglijabilă asupra caracteristicii dependente **y** are expresia:

Material cu drept de autor

$$\hat{y} = a + b.x \text{ sau } Y = a + b.x \qquad (7.91)$$

în care **a** şi **b** sunt coeficienți (parametri) ce urmează a fi calculați, iar **Y** se citeşte „**Y ajustat după x**".

Parametri **a** şi **b** au în acest caz conținut de medii şi se estimează cu ajutorul unor metode specifice oferite de statistică, ca de exemplu: metoda celor mai mici pătrate.

La folosirea metodei celor mai mici pătrate se presupune că suma pătratelor abaterilor dintre valorile empirice (reale) **y** şi valorile teoretice (ajustate) **Y** să fie minimă, adică:

$$\Sigma(y_i - Y)^2 = \min$$

respectiv:

$$\Sigma(y_i - a - b.x_i)^2 = \min$$

Derivând în raport cu **a** şi **b** şi anulând derivatele parțiale se obține sistemul de ecuații normale:

$$\begin{cases} na + b\Sigma x_i = \Sigma y_i \\ a\Sigma x_i + b.\Sigma x_i^2 = \Sigma x_i y_i \end{cases}$$

Rezolvând sistemul de ecuații normale se obțin formulele uzuale de calcul ale parametrilor ecuației de regresie:

$$a = \frac{\Delta\ a}{\Delta} = \frac{\begin{vmatrix} \Sigma y_i & \Sigma x_i \\ \Sigma x_i y_i & \Sigma x_i^2 \end{vmatrix}}{\begin{vmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{vmatrix}} = \frac{\Sigma y_i \times \Sigma x_i^2 - \Sigma x_i y_i \times \Sigma x_i}{n \times \Sigma x_i^2 - (\Sigma x_i)^2}$$

$$b = \frac{\Delta\ b}{\Delta} = \frac{\begin{vmatrix} n & \Sigma y_i \\ \Sigma x_i & \Sigma x_i y_i \end{vmatrix}}{\begin{vmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{vmatrix}} = \frac{n \times \Sigma x_i y_i - \Sigma x_i \times \Sigma y_i}{n \times \Sigma x_i^2 - (\Sigma x_i)^2}$$

Înlocuind valorile din tabelul 7.7 în ecuațiile normale rezultă:

$$\begin{cases} 5.a + 10.b = 26 \\ 10.a + 30.b = 67{,}5 \end{cases} \Rightarrow \begin{cases} a = 2{,}1 \\ b = 1{,}55 \Rightarrow \hat{y} = 2{,}1 + 1{,}55.x \\ \bar{y} = 5{,}2 \end{cases}$$

Material cu drept de autor

$$R = \sqrt{1 - \frac{0,275}{24,25}} = 0,994; \quad R^2 = 0,988;$$

$$R = \sqrt{1 - \frac{159,5 - 2,1 \times 26 - 1,55 \times 67,5}{159,5 - 26^2 / 5}} = 0,994$$

Dacă perechile de valori $\mathbf{x_i y_i}$ se întâlnesc, în cadrul distribuției, de mai multe ori **„$\mathbf{n_i}$"**, atunci cele două variabile vor apare ponderate în cadrul sistemului de ecuații normale:

$$\begin{cases} a\Sigma n_i + b\Sigma n_i x_i = \Sigma n_i y_i \\ a\Sigma n_i x_i + b.\Sigma n_i x_i^2 = \Sigma n_i x_i y_i \end{cases} \quad (7.92)$$

Rezolvând sistemul se obțin formulele de calcul ale celor doi parametri:

$$a = \frac{\Sigma n_i y_i \times \Sigma n_i x_i^2 - \Sigma n_i x_i y_i \times \Sigma n_i x_i}{\Sigma n_i \times \Sigma n_i x_i^2 - \left(\Sigma n_i x_i\right)^2} \quad (7.93)$$

$$b = \frac{\Sigma n_i \times \Sigma n_i x_i y_i - \Sigma n_i x_i \times \Sigma n_i y_i}{\Sigma n_i \times \Sigma n_i x_i^2 - \left(\Sigma n_i x_i\right)^2} \quad (7.94)$$

Exemplu de calcul 7.5. La 100 unități comerciale se analizează dependența dintre volumul vânzărilor şi numărul de salariați. Datele înregistrate se prezintă în tabelul 7.8.

Tabelul 7.8

Grupe de **S.C.** după nr. de salariați (**x**)	Subgrupe de **S.C.** după valoare vânzărilor (u.m.) (**y**)							Total
	14-18	18-22	22-26	26-30	30-34	34-38	38-42	
3-9	4	1						5
9-15		4	5					9
15-21		5	7	4	1			17
21-27			11	9	6	1		27
27-33			5	7	7	5		24
33-39				2	6	2	1	11
39-45						3	4	7
Total	4	10	28	22	20	11	5	100

Material cu drept de autor

Dispunerea pe diagonală a datelor înregistrate rezultă că între numărul de salariați şi volumul vânzărilor la **S**ocietățile **C**omerciale vizate este foarte intensă şi evident liniară: $Y_x = a + b.x$

Calculul valorilor funcției de regresie s-a făcut pe baza utilizării tabelului 7.9. A rezultat:

$$\begin{cases} a\Sigma\Sigma n_{ij} + b\Sigma x_i n_i = \Sigma y_j n_j \\ a\Sigma x_i n_i + b\Sigma x_i^2 n_i = \Sigma x_i y_j n_{ij} \end{cases} \Rightarrow \begin{cases} a.100 + b.2502 = 2788 \\ a.2502 + b.70740 = 73992 \end{cases}$$

a = 14,87896; b = 0,520425

$$Y_{x_i} = 14{,}88 + 0{,}52.x_i$$

Parametrul **a = 14,88** reprezintă ordonata la origine şi arată că nivelul mediu al vânzărilor ar fi fost de **14,88 u**nități **m**onetare dacă asupra acestora ar fi influențat alte cauze în afară de numărul de salariați. Parametrul **b = 0,52 u.m**., denumit şi coeficient de regresie, este pozitiv şi arată că dacă numărul de salariați ar fi crescut în medie cu o persoană, volumul mediu al vânzărilor ar fi crescu cu 0,52 u.m. Se observă că între valorile reale (empirice) y_j şi valorile ajustate Y_x ale variabilei rezultative există diferențe. Explicația acestor diferențe constă în aceea că valorile reale reflectă influența tuturor factorilor care au acționat asupra vânzărilor, pe când valorile (ajustate) calculate pe baza funcției de regresie reflectă numai tendința medie, legitatea care se manifestă în legătură cu numărul de salariați şi valoarea vânzărilor, deci exclude influența factorilor neînregistrați.

Tabelul 7.9

S.C. u.m. / nr. salariați		14-18	18-22	22-26	26-30	30-34	34-38	38-42	Total, n_i	$x_i.n_i$	$x_i^2.n_i$	$x_i\Sigma y_j n_{ij}$
	y_j / x_i	16	20	24	28	32	36	40				
3-9	6	4	1						5	30	180	504
9-15	12		4	5					9	108	1296	2400
15-21	18		5	7	4	1			17	306	5508	7416
21-27	24			11	9	6	1		27	648	15552	17856
27-33	30			5	7	7	5		24	720	21600	21600
33-39	36				2	6	2	1	11	396	14256	12960
39-45	42						3	4	7	294	12348	11256
Total n_j		4	10	28	22	20	11	5	$\Sigma n_i = \Sigma n_j$ 100	$\Sigma x_i n_i$ 2502	$\Sigma x_i^2 n_i$ 70740	$\Sigma x_i y_j n_{ij}$ 73992
$y_j n_j$		64	200	672	616	640	396	200	$\Sigma y_j n_j$ 2788			
$y_j^2 n_j$		1024	4000	16128	17248	20480	14256	8000	$\Sigma y_j^2 n_j$ **81136**			

Material cu drept de autor

$y_j\Sigma x_i n_{ij}$	384	2880	14400	15960	18816	13392	8160	$\Sigma y_j x_i n_{ij}$ **73992**
$Y_x=a+bx$	17,98	21,1	24,23	27,35	30,47	33,59	36,72	-

Intensitatea legăturii se poate determina prin coeficientul de corelație $\mathbf{r_{y/x}}$ şi raportul de corelație $\mathbf{R_{y/x}}$.

$$r_{y/x} = \frac{\Sigma\Sigma n_{ij}.\Sigma x_i y_j n_{ij} - \Sigma x_i n_i.\Sigma y_j n_j}{\sqrt{[\Sigma n_i.\Sigma x_i^2 n_i - (\Sigma x_i n_i)^2]\times[\Sigma n_j.\Sigma y_j^2 n_j - (\Sigma y_j n_j)^2]}} =$$

$$= \frac{100\times 73992 - 2502\times 2788}{\sqrt{[100\times 70740 - 2502^2]\times[100\times 81136 - 2788^2]}} = 0{,}8$$

$$R_{y/x} = \sqrt{1 - \frac{\Sigma y_j^2 n_j - a\Sigma y_j n_j - b\Sigma x_i y_j n_{ij}}{\Sigma y_j^2 n_j - \frac{(\Sigma y_j n_j)^2}{\Sigma n_j}}} = \sqrt{1 - \frac{81136 - 14{,}88.2788 - 0{,}52.73992}{81136 - \frac{2788^2}{100}}} = 0{,}8$$

Se observă că $\mathbf{r_{y/x} = R_{y/x}}$ **(0,8 = 0,8)** ceea ce înseamnă că legătura dintre numărul de salariați şi valoarea vânzărilor de mărfuri este directă şi destul de intensă (puternică) şi de formă liniară.

Dacă se calculează coeficientul de determinație $\mathbf{R^2_{y/x} = (R_{y/x})^2 = 0{,}8^2 = 0{,}64}$ sau $\mathbf{R^2_{y/x} = 64\%}$ aceasta înseamnă că influența exercitată de creşterea numărului de salariați influențează pozitiv volumul vânzărilor numai în proporție de **64%**, restul de **36%** se datorează altor cauze, considerate constante sau cu caracter întâmplător, care nu au fost luate în considerare în modelul unifactorial de mai sus.

Se poate testa şi semnificația coeficientului de corelație cu testul t a lui Student, folosindu-se relația:

$$t = \frac{r_{y/x}}{\sqrt{1 - r_{y/x}^2}}\sqrt{n-2} \qquad (7.95)$$

Pe baza expresiei (7.95) se obține:

$$t = \frac{r_{y/x}}{\sqrt{1 - r_{y/x}^2}}\sqrt{n-2} = \frac{0{,}8}{\sqrt{1 - 0{,}8^2}}\sqrt{100-2} = 13{,}19$$

Din anexa 3, $\mathbf{t_{\alpha=0,05;n=n-2=98\ grade\ de\ libertate} = 1{,}984}$. Întrucât $\mathbf{t_{calculat} = 13{,}19 > t_{tabelat} = 1{,}984}$ se poate admite că coeficientul de corelație este semnificativ, respectiv legătura dintre numărul de salariați şi valoarea vânzărilor nu este întâmplătoare.

7.9. Regresia neliniară

La analiza de regresie se întâlnesc cazuri când în intervalul cercetat relația dintre două variabile nu poate fi exprimată printr-o dreaptă.

Material cu drept de autor

În practică se deosebesc două situații şi în consecință două procedee pentru exprimarea relației de dependență neliniară între două variabile.

a) relația dintre variabilele **x** şi **y** nu este liniară; ea însă poate fi transformată în aşa fel, încât să devină liniară. În acest caz coeficienții de regresie se pot calcula uşor prin metoda cunoscută de la analiza regresiei liniare.

b) nu există nici o posibilitate simplă de transformare a relației neliniare într-o relație liniară. În acest caz, din grafica regresiei se apreciază forma relației (parabolică, hiperbolică etc.) calculând apoi coeficienții de regresie ai funcției cu ajutorul ecuațiilor normale corespunzătoare.

Pentru liniarizarea curbelor vom considera drept exemplu caracteristic ecuația:

$$\mathbf{Y = k.e^{c.x}} \tag{7.96}$$

care prin logaritmare conduce la relația liniară:

$$\mathbf{\log Y = \log k + 0{,}4343.c.x} \tag{7.97}$$

în care **a = log k** şi **b = 0,4343.c**

Un alt exemplu este relația:

$$\mathbf{Y = k.x^{b}} \tag{7.98}$$

care, prin logaritmare, se transformă în dreapta:

$$\mathbf{\log Y = \log k + b.\log x} \tag{7.99}$$

pentru care coeficienții de regresie **a = log k** şi **b** se calcula cu ajutorul metodei descrise pentru regresia liniară cu deosebirea că în schema de calcul valorilor **y** şi **x**, valorile logaritmice ale acestora şi se va reveni apoi la valorile normale cu ajutorul antilogaritmilor.

Există însă situații când relația între variabile nu poate fi liniarizată. Un exemplu tipic este relația parabolică de gradul doi:

$$\mathbf{Y = b_0 + b_{1.x} + b_2.x^2} \tag{7.100}$$

sau un polinom de grad superior. Pentru calculul coeficienților de regresie $\mathbf{b_0}$, $\mathbf{b_1}$, $\mathbf{b_2}$,... se va folosi şi în acest caz metoda celor mai mici pătrate. Va trebui deci minimizată expresia valabilă pentru cazul general:

$$\sum_{i=1}^{n}\left[y_1 - f\left(x_i, b_0, b_1, b_2, \ldots\right)\right]^2 = \min \tag{7.101}$$

sau, în cazul de față:

$$\sum_{i=1}^{n}\left(y - Y\right)^2 = \sum_{i=1}^{n}\left[y_i - \left(b_0 + b_1.x_i + b_2.x_i^2\right)\right]^2 = \min \tag{7.102}$$

Egalând cu zero derivatele parțiale în raport cu coeficienții de regresie $\mathbf{b_0}$, $\mathbf{b_1}$ şi $\mathbf{b_2}$, după aranjarea termenilor se obține:

$$\left.\begin{aligned} n.b_0 + b_1\Sigma x_i + b_2\Sigma x_i^2 &= \Sigma y_i \\ b_0\Sigma x_i + b_1\Sigma x_i^2 + b_2\Sigma x_i^3 &= \Sigma y_i x_i \\ b_0\Sigma x_i^2 + b_1\Sigma x_i^3 + b_2 x_i^4 &= \Sigma y_i x_i^2 \end{aligned}\right\} \tag{7.103}$$

Material cu drept de autor

Prin rezolvarea sistemului de ecuații se obțin coeficienții de regresie ai funcției parabolice. O dată cu creşterea gradului polinomului, calculele devin mai complicate, fiind necesare puteri ale lui **x** de două ori mai mari decât gradul polinomului ales. Astfel, de exemplu, pentru cazul unui polinom de gradul trei trebuie rezolvat un sistem de ecuații cu patru necunoscute pentru care trebuie calculate sumele puterilor $\mathbf{x^4}$, $\mathbf{x^5}$ şi $\mathbf{x^6}$ însă de cele mai multe ori în analiza fenomenelor economico-sociale funcțiile de gradul doi se vor descrie fenomenul cercetat cu o exactitate satisfăcătoare. este necesar să se evite construirea curbelor empirice, care contrazic legi sau relații cunoscute sau care nu se pot justifica din punct de vedere fizic.

Exemplu de calcul 7.6. Se consideră seria statistică de valori pentru variabila independentă **X** şi variabila dependentă **Y**:

x_i	5	6	10	12	15	20	22	30	35	40
y_i	70	53	44	38	35	28	26	17	11	5

Se cere:

a. Să se analizeze existența, direcția şi forma legăturii dintre variabilele **X** şi **Y**.

b. Să se determine parametrii funcției de regresie.

c. Să se precizeze intensitatea legăturii dintre **X** şi **Y**.

Rezolvare.

a. Se alcătuieşte corelograma din figura următoare şi se observă că între cele două variabile există o legătură inversă, neliniară.

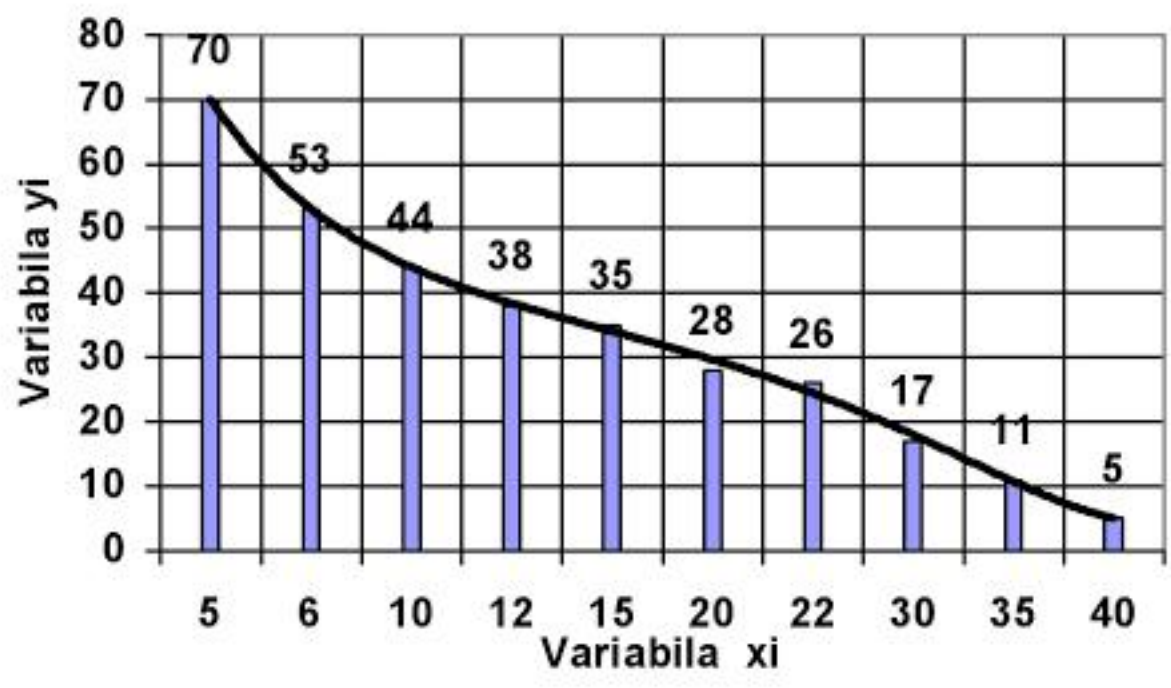


Fig.7.3. Reprezentarea grafică a datelor problemei din exemplul 7.5.

b. pentru modelarea tendinței legăturii, se va folosi modelul de regresie hiperbolic:

$$Y_{x_i} = a + b\frac{1}{x_i}$$

Aplicând metoda celor mai mici pătrate se obține sistemul:

Material cu drept de autor

$$\begin{cases} na + b\Sigma\frac{1}{x_i} = \Sigma y_i \\ a\Sigma\frac{1}{x_i} + b\Sigma\frac{1}{x_i^2} = \Sigma\frac{1}{x_i}y_i \end{cases}$$

$$a = \frac{\begin{vmatrix} \Sigma y_i & \Sigma\frac{1}{x_i} \\ \Sigma\frac{1}{x_i}y_i & \Sigma\frac{1}{x_i^2} \end{vmatrix}}{\begin{vmatrix} n & \Sigma\frac{1}{x_i} \\ \Sigma\frac{1}{x_i} & \Sigma\frac{1}{x_i^2} \end{vmatrix}} = \frac{\Sigma y_i \times \Sigma\frac{1}{x_i^2} - \Sigma\frac{1}{x_i} \times \Sigma\frac{1}{x_i}y_i}{n \times \Sigma\frac{1}{x_i^2} - \left(\Sigma\frac{1}{x_i}\right)^2} =$$

$$= \frac{327 \times 0{,}0938 - 0{,}791 \times 35{,}943}{10 \times 0{,}0938 - 0{,}6257} = 7{,}18$$

$$b = \frac{\begin{vmatrix} n & \Sigma y_i \\ \Sigma\frac{1}{x_i} & \Sigma\frac{1}{x_i}y_i \end{vmatrix}}{\begin{vmatrix} n & \Sigma\frac{1}{x_i} \\ \Sigma\frac{1}{x_i} & \Sigma\frac{1}{x_i^2} \end{vmatrix}} = \frac{n \times \Sigma\frac{1}{x_i}y_i - \Sigma\frac{1}{x_i} \times \Sigma y_i}{n \times \Sigma\frac{1}{x_i^2} - \left(\Sigma\frac{1}{x_i}\right)^2} =$$

$$= \frac{10 \times 35943 - 0{,}791 \times 327}{10 \times 0{,}0938 - 0{,}6257} = 322{,}68$$

$$Y_{x_i} = 7{,}18 + 322{,}68 \times \frac{1}{x_i}$$

Valorile ajustate ale lui **Y** şi calculele intermediare se prezintă în tabelul 7.9.

Material cu drept de autor

Tabelul 7.9

x_i	y_i	$1/x_i$	$y_i(1/x_i)$	$(1/x_i)^2$	Y_{x_i}	$(y_i - Y_{x_i})^2$	$(y_i - \bar{y})^2$
0	**1**	**2**	**3**	**4**	**5**	**6**	**7**
5	70	0,200	14,000	0,0400	71,58	2,49	1391,29
6	53	0,160	8,480	0,0256	58,70	32,49	412,09
10	44	0,100	4,400	0,0100	39,38	21,34	127,69
12	38	0,083	3,154	0,0069	33,90	16,81	28,09
15	35	0,067	2,345	0,0044	28,75	39,01	5,29
20	28	0,050	1,400	0,0025	23,28	22,28	22,09
22	26	0,045	1,170	0,0020	21,67	18,75	44,89
30	17	0,033	0,561	0,0010	17,80	0,64	246,49
35	11	0,028	0,308	0,0008	16,20	27,04	470,89
40	5	0,025	0,125	0,0006	15,23	104,65	767,29
-	327	0,791	35,943	0,0938		285,50	3516,10

c. Intensitatea legăturii se stabileşte cu raportul de corelație:

$$R_{y/x} = \sqrt{1 - \frac{\Sigma(y_i - Y_{x_i})^2}{\Sigma(y_i - \bar{y})^2}} = \sqrt{1 - \frac{285,5}{3516,1}} = 0,96$$

$$\bar{y} = \frac{\Sigma y_i}{n} = \frac{327}{10} = 32,7$$

Raportul de corelație arată o legătură puternică între cele două variabile.

Coeficientul de determinație $R^2_{y/x} = 0,92$ arată că 92% din variația variabilei **Y** este explicată prin influența lui **X** asupra lui **Y**.

7.10. Corelația neparametrică

Metodele de studiere a corelației dintre fenomene, prezentate până acum se aplică în practică numai în cazurile când se lucrează cu caracteristici cantitative distribuite după legea normală, construite pe scale de intervale sau de rapoarte. Aceste metode poartă denumirea de metode parametrice. Dacă distribuirea variabilelor corelate nu este normală, iar acestea sunt apreciate calitativ, prin atribute, pentru măsurarea intensității corelației dintre variabile se vor utiliza metode neparametrice. Caracteristicile calitative vor fi considerate mai întâi **dihotomice**, apoi ca fiind construite pe scale nominale (variabile categoriale) şi, în final, ca variabile ordinale respectându-se principiul general conform căruia procedeul valabil pentru o scală inferioară este valabil şi pentru una superioară.

Material cu drept de autor

O caracteristică dihotomică este aceea care realizează o clasificare a subiecților în două categorii, adică o variabilă de tipul „da/nu", „peste/sub", adică posedă sau nu o anumită proprietate. Se mai numeşte şi atributivă, tocmai pentru că face distincția între subiecții care posedă un anumit atribut şi cei care nu-l posedă. Deci, metodele neparametrice se utilizează atunci când cel puțin una dintre variabile este calitativă sau când variabilele nu au o distribuție normală sau asimptotic normală. Metodele neparametrice au fost elaborate pentru prima dată de către Spearman şi au fost dezvoltate de către K. Pearson, U. Yulle, M. Kendall şi alți statisticieni.

Coeficienții corelației neparametrice se bazează:

- pe frecvențele $\mathbf{f_{ij}}$ ale perechilor de valori $\mathbf{(x_i,y_j)}$ şi pe frecvențele marginale dintr-un tabel de corelație, fie;
- pe regula de adunare a dispersiilor aplicabilă pe tabelul de corelație, fie;
- pe rangurile ce se acordă unităților statistice în raport cu fiecare din caracteristicile luate în studiu.

Coeficienții corelației neparametrice sunt aplicabili atât în cazul legăturilor dintre caracteristici numerice cât şi în cel al caracteristicilor nenumerice.

Cele mai obişnuite aplicații în cercetarea legăturilor statistice în acest sens sunt:

- asocierea caracteristicilor dihotomice;
- testul χ^2 al lui Pearson de verificare a existenței legăturii;
- coeficientul de contingență al lui Pearson;
- coeficienții simpli de corelație a rangurilor ai lui Kendall şi Spearmann;
- coeficientul lui Fechner (pentru concordanța semnelor) ş.a.

Baza analizei asocierii o constituie tabelul de asociere/ contingență. Pentru variabilele atributive tabelul are forma:

	Da (+)	Nu (-)	
Da (+)	a	b	a + b
Nu (-)	c	d	c + d
	a + c	b + d	

Ideea de bază este aceea că, **dacă subiecții se concentrează pe diagonala principală, deci dacă semnele coincid (+ + şi - -), între cele două variabile există o asociere**.

(a). Coeficientul Φ:

$$\Phi = \frac{n.a-(a+b)(a+c)}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \qquad (7.104)$$

Material cu drept de autor

unde numitorul cuprinde cele patru totaluri, iar n este numărul total al subiecților.

(b) Coeficientul de asociere **Q**:

$$Q = \frac{a.d - b.c}{a.d + b.c} \tag{7.105}$$

(c) Coeficientul de concordanță **Cc**:

$$Cc = \frac{(a+d) - (b+c)}{(a+d) + (b+c)} \tag{7.106}$$

Toți cei trei coeficienți de asociere iau valori între (-1) şi (+1) şi prezintă dezavantajul că nu iau în calcul mărimea abaterilor, ci doar semnele acestora, cu observația că întotdeauna **Q** >Φ. Din această cauză, ei sunt mai puțin recomandabili pentru alte variabile decât cele dihotomice.

Exemplu de calcul 7.7. Dintre cei 80 de studenți care au promovat primul test, 60 l-au promovat şi pe al doilea, în timp ce dintre cei 40 care nu au promovat primul test, 30 nu l-au promovat nici pe al doilea:

	Da (+)	Nu (-)	
Da (+)	60	20	80
Nu (-)	10	30	40
	70	50	

$$\Phi = \frac{120.60 - (60+20)(60+10)}{\sqrt{(60+20)(10+30)(60+10)(20+30)}} = 0{,}478$$

$$Q = \frac{60.30 - 20.10}{60.30 + 20.10} = 0{,}8$$

$$Cc = \frac{(60+30) - (20+30)}{(60+30) + (20+10)} = 0{,}5$$

Testul χ^2 al lui Pearson. În acest scop se realizează tabelul de contingență în care se sistematizează datele care au **r** rânduri (**r** clase pentru variabila **X**) şi **c** coloane (**c** clase pentru variabila **Y**), tabelul 7.10.

Pentru testul χ^2 de independență pentru tabelul „**r** x **c**" de contingență (asociere) se aplică sub presupunerea că fiecare observație (unitate statistică) este clasificată independent de orice altă observație. Ipoteza nulă constă în independența clasificării pe linie față de clasificarea pe coloane. Se vor determina atunci frecvențele teoretice (aşteptate) în rândul **i** şi coloana **j**.

$$f_{ij} = \frac{n_{i\bullet} \times n_{\bullet j}}{n_{\bullet\bullet}} \tag{7.107}$$

şi se va calcula testul statistic.

Material cu drept de autor

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij} - f_{ij})^2}{f_{ij}} \qquad (7.108)$$

Tabelul 7.10

Tabel de contingență

X \ Y	Y_1	Y_2	...	Y_j	...	Y_c	Total
X_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1c}	$n_{1\bullet}$
X_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2c}	$n_{2\bullet}$
...	...	...	...	...	...	...	...
X_i	n_{i1}	n_{i2}	...	n_{ij}	...	m_{ic}	
...	...	...	...	...	...	...	...
X_r	n_{r1}	n_{r2}	...	n_{rj}	...	n_{rc}	$n_{r\bullet}$
Total	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet j}$	...	$n_{\bullet c}$	$n_{\bullet\bullet}$

Ipoteza nulă se respinge (şi deci se acceptă ipoteza alternativă, aceea că există dependență între clasificarea pe linii şi cea pe coloane), la un nivel de semnificație α, dacă $\chi^2_{calc} > \chi^2_{\alpha,(r-1)(c-1)}$, unde (r–1)(c-1) reprezintă numărul gradelor de libertate.

Se apreciază că legătura este cu atât mai intensă cu cât distanța dintre cele două variabile χ^2 este mai mare.

Exemplu de calcul 7.8. S-a efectuat un sondaj pe un eşantion de 1200 consumatori privind analiza cererii pentru un anumit produs şi modul de apreciere a acestuia sub influența diferiților factori. Un segment din prelucrarea datelor este evidențiat în tabelul următor 7.11.

Se cere să se stabilească dacă vârsta influențează semnificativ aprecierea produsului alegând un nivel de semnificație α = **0,05**.

Rezolvare. Se va folosi criteriul χ^2. Se emite ipoteza nulă $\mathbf{H_0}$ conform căreia aprecierea totală (favorabil + nefavorabil) are aceleaşi proporții în cadrul fiecărei subgrupe de vârstă. În acest caz frecvențele în fiecare subgrupă ar trebui să fie cele din tabelul 7.12.

Tabelul 7.11

Aprecieri	Grupe de vârstă (ani)			Total
	Sub 30	30 - 50	Peste 50	
Favorabile	193	232	139	564
Nefavorabile	230	241	165	636
Total	423	473	304	1200

Material cu drept de autor

Tabelul 7.12

Aprecieri	Grupe de vârstă (ani)			Total
	Sub 30	30 - 50	Peste 50	
Favorabile	564.423/1200≈ 199 (47%.423)	564.473/1200≈ 222 (47%.473)	564.304/1200≈ 143 (47%.304)	564 (47%)
Nefavorabile	636.423/1200≈ 224 (53%.423)	636.473/1200≈ 251 (53%.473)	636.304/1200≈ 161 (53%.304)	636 (53%)
Total	423 (100%)	473 (100%)	304 (100%)	1200 (100%)

Pe baza tabelului de mai sus se calculează mărimea χ^2. Mărimea χ^2 exprimă distanța dintre cele două distribuții de frecvențe (distribuția empirică $\mathbf{n}_{ij}$ şi distribuția teoretică $\mathbf{f}_{ij}$) şi se calculează însumând rapoartele dintre pătratele diferențelor elementelor celor două distribuții, empirice şi teoretice.

În sine, mărimea χ^2 nu măsoară asocierea între cele două distribuții, ci oferă informații asupra relației dintre cele două variabile care au determinat constituirea distribuțiilor. Valoarea χ^2 astfel determinată, comparată cu valoarea tabelară pentru **(nr.linii-1)(nr.coloane-1)** grade de libertate, arată dacă variabilele studiate sunt sau nu independente.

$$\chi^2 = \frac{(193-199)^2}{199} + \frac{(232-222)^2}{222} + \frac{(139-143)^2}{143} +$$

$$+ \frac{(230-224)^2}{224} + \frac{(421-251)^2}{251} + \frac{(165-161)^2}{161} = 1{,}401$$

HI – pătrat calculat se compară cu HI – pătrat tabelat.

Pentru $\alpha = 0{,}05$ şi $\nu = (2-1)(3-1) = 2$, $\chi^2 = 5{,}991$.

Deoarece:

$$\chi^2_{calculat} = 1{,}401 < \chi^2_{tab} = 5{,}991$$

ipoteza nulă se admite, adică aprecierile persoanelor privind produsul nu sunt influențate de vârstă.

Coeficientul de asociere a lui Pearson:

$$C_{Pearson} = \sqrt{\frac{\chi^2}{n+\chi^2}} \qquad (7.109)$$

şi ia valori între 0 şi 1: apropierea de zero înseamnă legătură slabă, iar apropierea de unu înseamnă legătură puternică.

Material cu drept de autor

În exemplul de mai sus:

$$C_{Pearson} = \sqrt{\frac{1,401}{1200+1,401}} = 0,0012$$

confirmă faptul că între vârstă şi aprecierea persoanelor privind produsul nu există legătură.

Întrucât în calculul variabilei χ^2 şi al coeficientului $C_{Pearson}$ se folosesc doar frecvenţele $\mathbf{f_{ij}}$ şi frecvenţele marginale $\mathbf{f_i}$ şi $\mathbf{f_j}$, nu şi variantele celor două caracteristici, aceste două metode se pot folosi şi în cazul analizei legăturilor dintre variabile nenumerice.

Prezintă ca deficienţă faptul că valoarea lui nu poate fi niciodată egală cu unitatea (numitorul fiind totdeauna mai mare decât numărătorul). Pe măsură ce creşte dimensiunea tabelului, limita maximă a coeficientului tinde către unitate. Pentru un tabel pătratic, cu structurile marginale coincidente şi cu toţi subiecţii pe diagonală, valorile maxime sunt: 0,707 pentru un tabel cu 2 linii şi 2 coloane; 0,816 pentru un tabel cu 3 linii şi 3 coloane; 0,866 pentru un tabel cu 4 linii şi 4 coloane etc.

Coeficientul de contingenţă al lui Pearson. În construirea acestui indicator autorul porneşte de la regula de adunare a dispersiilor, adică varianţa totală a lui **y** este constituită din varianţa datorată influenţei lui **x** (varianţa explicită) + varianţa datorată altor factori (varianţa neexplicită sau reziduală), respectiv:

$$\sigma^2 = \sigma^2_{y/x} + \overline{\sigma}^2 \qquad (7.110)$$

unde σ^2 este dispersia totală; $\sigma^2_{y/x}$ este dispersia dintre grupe; $\overline{\sigma}^2$ este dispersia medie.

Pearson a definit intensitatea legăturii dintre variabilele **y** şi **x** în funcţie de ponderea pe care o deţine varianţa lui **y** datorită influenţei lui **x** în varianţa totală a lui **y**. Rădăcina pătrată din această pondere se numeşte raportul de corelaţie al lui Pearson şi se noatează:

η sau $\mathbf{R_{y/x}}$ când se lucrează pe o populaţie statistică întreagă;

η' sau $\mathbf{R'_{y/x}}$ când se lucrează pe eşantion.

$$\eta = \sqrt{\frac{\sigma^2_{y/x}}{\sigma^2}} = \sqrt{1 - \frac{\overline{\sigma}^2}{\sigma^2}} \qquad (7.111)$$

În literatura de specialitate, unii autori adoptă formula pentru coeficientul corelaţiei parametrice neliniare.

Când raportul de corelaţie calculat cu expresia de mai sus se calculează pe un eşantion se pune întrebarea în ce măsură intensitatea legăturii la nivelul întregii populaţii este aceeaşi ? Adică, în ce măsură raportul de corelaţie este semnificativ ?

Material cu drept de autor

Pentru verificarea semnificației raportului de corelație se folosește „**Testul F al lui Fisher Snedecor**".

Se emite ipoteza nulă: H_0: η este nesemnificativ; H_1: η este semnificativ.

Se construieşte variabila **F** ca raport a două dispersii, astfel:

$$F_{calculat} = \frac{s_1^2}{s_2^2} \qquad (7.112)$$

unde s_1^2 este estimația dispersiei între grupe în populația totală la $\nu_1 = m - 1$ grade de libertate şi s_2^2 estimația dispersiei în interiorul grupelor în populația totală la $\nu_2 = n - m$ grade de libertate.

Conținutul şi modul de calcul ale elementelor necesare aplicării testului **F** sunt prezentate în tabelul 7.13.

Raportul $F_{calculat} = \frac{s_1^2}{s_2^2}$ devine

$$F_{calculat} = \frac{m-n}{n-1} \times \frac{\eta^2}{1-\eta^2}$$

Valoarea calculată pentru **F** se compară cu o valoare tabelată corespunzătoare unui grad de semnificație ales, de obicei 0,05 şi gradelor de libertate ν_1 şi ν_2 (aflate pe orizontala şi respectiv pe verticala tabelului).

Dacă $F_{calculat} < F_{tabelat}$, ipoteza nulă H_0 se respinge, respectiv η este semnificativ pentru intensitatea legăturii dintre cele două variabile, concluzia se poate extinde la nivelul întregii populații statistice.

Tabelul 7.13

Felul variației	Numărul gradelor de libertate	Estimația dispersiei
Totală	$\nu = m - 1$	$s_0^2 = \frac{\sigma^2}{m-1}$
Între grupe	$\nu_1 = n - 1$	$s_1^2 = \frac{\sigma_{y/x}^2}{n-1}$
În interiorul grupelor	$\nu_2 = m - n$	$s_2^2 = \frac{\overline{\sigma}^2}{m-n}$

Material cu drept de autor

Dacă $F_{calculat} > F_{tabelat}$, ipoteza nulă H_0 se admite, respectiv η **nu** este semnificativ pentru intensitatea legăturii dintre cele două variabile, concluzia nu se poate extinde la nivelul întregii populații statistice.

Coeficienți simpli de corelație a rangurilor.

Coeficientul Kendall. În construirea acestui indicator autorul pornește de la o populație statistică cu privire la care se dorește analiza corelației dintre variabilele **x** şi **y**.

Se consideră cazul cel mai simplu, când nici o variantă nu se repetă şi în acest caz, distribuind unitățile pe cele două variabile se obține:

x: $x_1, x_2, \ldots, x_i, \ldots, x_n$.

y: $y_1, y_2, \ldots, y_j, \ldots, y_n$.

Se ordonează rangurile după variabila factorială **x** şi se înscriu, în paralel, pe şirul al doilea, rangurile acordate după cea de a doua variabilă rezultativă **y**. Se notează apoi cu p_i numărul de ranguri superioare rangului **i** al variabilei dependente care există pe coloana respectivă după fiecare rang, iar cu q_i numărul de ranguri inferioare variabilei dependente care există după fiecare rang al acesteia.

Coeficientul lui Kendall se exprimă cu relația:

$$k = \frac{2(\Sigma p_i - \Sigma q_i)}{n.(n-1)} = \frac{2.(P-Q)}{n.(n-1)} = \frac{2.S}{n.(n-1)} \qquad (7.113)$$

$$S = P - Q; \quad P = \Sigma p_i; \quad Q = \Sigma q_i$$

Între rangurile perechi poate exista concordanță totală, discordanță totală sau o situație intermediară. De aceea, coeficientul lui Kendall variază tot între **–1** şi **+1**. Astfel, dacă între rangurile celor două variabile există o concordanță deplină, atunci **Q = 0**, iar **k = 1**; dacă între ranguri există o discordanță totală atunci **P = 0** şi **k = -1**.

Semnificația coeficientului este următoarea. Semnul lui **k** arată sensul legăturii: când **k > 0** (adică **P > 0**) între variabile există legătură directă; când **k < 0** (adică **P < 0**) între variabile există legătură inversă. Când **k** tinde la **1** (adică **P** şi **Q** tind spre valoarea maximă) legătura dintre cele două variabile este mai intensă. Când **k** tinde la **0** (adică **P** şi **Q** sunt apropiate) legătura dintre cele două variabile este mai slabă.

Coeficientul Spearman. Caracterizează intensitatea legăturii dintre caracteristica factorială **x** şi caracteristica rezultativă **y** în ipoteza existenței unei legături între cele două variabile.

Autorul pornește de la expresia:

$$r_{y/x} = \frac{\Sigma(x_i - \bar{x}).(y_j - \bar{y})}{n.\sigma_x.\sigma_y} \qquad (7.114)$$

reprezentând coeficientul de corelație simplă liniară.

Se fac notațiile: $X = x - \bar{x}$ şi $Y = y - \bar{y}$ cărora li se asociază mărimile lor. De remarcat faptul că rangurile aferente diferențelor sunt aceleaşi cu rangurile variantelor.

Material cu drept de autor

Considerând că rangurile variabilelor **X** şi **Y** sunt constituite din primele n numere naturale, întregi şi pozitive, nivelul mediu al celor două variabile va fi dat de expresiile:

$$\bar{x} = \bar{y} = \frac{n+1}{2} \tag{7.115}$$

iar abaterile medii pătratice se vor calcula cu expresiile:

$$\sigma_x = \sqrt{\frac{\Sigma x_i^2}{n} - \bar{x}^2} \; ; \quad \sigma_y = \sqrt{\frac{\Sigma y_j^2}{n} - \bar{y}^2} \tag{7.116}$$

Suma pătratelor primelor n numere naturale este dată de expresia:

$$1^2 + 2^2 + 3^2 + ... + n^2 = \frac{n.(n-1).(2.n+1)}{6} = \Sigma x_i^2 = \Sigma y_i^2$$

$$\sigma_x = \sigma_y = \sqrt{\frac{\frac{n.(n+1).(2.n+1)}{6}}{n} - \left(\frac{n+1}{2}\right)^2} = \sqrt{\frac{n^2-1}{12}} \tag{7.117}$$

$$n.\sigma_x.\sigma_y = n\frac{n^2-1}{12}$$

Se introduce ideea de distanță „**d**" reprezentând diferența dintre rangurile perechi ale variabilelor analizate, astfel: $\mathbf{d_i = x_i - y_j}$. Va rezulta:

$$\Sigma d_i^2 = \frac{2.n.(n^2-1)}{12} - 2.\Sigma(x_i - \bar{x}).(y_j - \bar{y})$$

de unde rezultă:

$$\Sigma(x_i - \bar{x}).(y_j - \bar{y}) = \frac{n.(n^2-1)}{12} - \frac{6.\Sigma d_i^2}{12}$$

Substituind în expresia coeficientului de corelație Pearson, rezultă coeficientul lui Spearman:

$$c = 1 - \frac{6\Sigma d^2}{n.(n^2-1)} \tag{7.118}$$

Semnificația coeficientului lui Spearman este similară coeficientului lui Kendall.

Exemplu de calcul 7.9. Se cunosc următoarele date cu privire la capitalul utilizat şi volumul vânzărilor dintr-o lună, pentru zece firme, principale producătoare ale unei game de produse, tabelul 7.14.

Material cu drept de autor

Tabelul 7.14

Firma	Capital utilizat (u.m.)	Vânzări (u.m.)
F_1	20,46	75,6
F_2	13,36	35,7
F_3	25,31	104,9
F_4	33,73	129,6
F_5	25,40	71,8
F_6	35,82	179,7
F_7	13,35	53,7
F_8	22,76	65,3
F_9	21,23	74,3
F_{10}	19,28	55,3

Să se stabilească dacă între cele două variabile există legătură, de ce sens şi de ce intensitate.

Rezolvare. Pentru a calcula coeficienții Kendall şi Spearman se ordonează datele într-un tabel de forma celui din 7.15.

Coeficientul Kendal:

$$k = \frac{2.(P-Q)}{n.(n-1)} = \frac{2.30}{10.(10-1)} = 0{,}67 \; sau \; 67\%$$

Coeficientul Spearman:

$$c = 1 - \frac{6\Sigma d^2}{n.(n^2-1)} = \frac{6.26}{10.(100-1)} = 0{,}84 \; sau \; 84\%$$

Se apreciază că între caracteristica „volumul capitalului utilizat" şi caracteristica „volumul vânzărilor" există o legătură destul de puternică.

Tabelul 7.15

Firma	Rangul după:		d	d^2	P	Q	P - Q
	x	y					
F_6	1	1	0	0	9	0	9
F_4	2	2	0	0	8	0	8
F_5	3	6	-3	9	4	3	1
F_3	4	3	1	1	5	0	5
F_8	5	7	-2	4	3	2	1
F_9	6	5	1	1	3	1	2
F_1	7	4	3	9	3	0	3
F_{10}	8	8	0	0	2	0	2
F_2	9	10	-1	1	0	1	-1
F_7	10	9	1	1	0	0	0
Total	-	-	0	26	37	7	30

Material cu drept de autor

Coeficientul rangurilor calculat după formula lui Kendall este, de obicei, mai mic decât cel calculat după formula lui Spearman.

Coeficientul lui Fechner. Presupune determinarea unui coeficient de concordanță denumit **coeficient simplu de covariație diferențială**:

$$K = \frac{c-d}{n}, \qquad K \in [-1,+1] \tag{7.119}$$

unde c şi d reprezintă numărul de concordanțe, respectiv discordanțe de semne ale abaterilor.

$$\Delta x_i = x_i - x_{i-1};\ \Delta y_j = y_j - y_{j-1} \text{ sau}$$

$$\Delta x_i = x_i - \bar{x};\ \Delta y_j = y_j - \bar{y}$$

1. Dacă există numai concordanțe de semne:

$$\Rightarrow \begin{cases} d = n \\ c = 0 \end{cases} \Rightarrow K = \frac{c-d}{n} = \frac{n-0}{n} = 1$$

2. Dacă există numai discordanțe de semne:

$$\Rightarrow \begin{cases} d = n \\ c = 0 \end{cases} \Rightarrow K = \frac{c-d}{n} = \frac{0-n}{n} = -1$$

3. Dacă numărul concordanțelor este egal cu numărul de discordanțe, atunci:

$$c = d \Rightarrow K = \frac{c-d}{n} = 0$$

Coeficientul lui Fechner are dezavantajul că ține seama doar de semnele abaterilor şi nu de mărimea acestor abateri Δx_i, Δy_i, de aceea se calculează un **coeficient ponderat de concordanță**.

$$K = \frac{C-D}{C+D} \tag{7.120}$$

unde, $C = \Sigma\Delta x_i.\Delta y_j$, pentru $\Delta x_i.\ \Delta y_j > 0$; $D = \Sigma\Delta x_i.\Delta y_j$, pentru $\Delta x_i.\Delta y_j < 0$; $K \in [-1,+1]$.

Exemplu de calcul 7.10. Se cunosc următoarele date privind salariul negociat şi vechimea în muncă a 10 angajați ai unei firme:

Tabelul 7.16

Nr. crt	Vechimea în muncă (ani)	Salariul negociat (u.m.)
1	5	600
2	6	500
3	8	900
4	10	1300

Material cu drept de autor

5	13	1200
6	15	1400
7	17	1600
8	18	1800
9	19	2000
10	20	1900

Să se determine în ce măsură cele două variabile se corelează.

Rezolvare. Se va folosi metoda lui Fechner.

Tabelul 7.17

Nr. crt	Vechimea în muncă (ani)	Salariul negociat (u.m.)	Δx_i	Δy_j	**C** sau **D**
1	5	600	-	-	-
2	6	500	1	-100	D
3	8	900	2	400	C
4	10	1300	2	400	C
5	13	1200	3	-100	D
6	15	1400	2	200	C
7	17	1600	2	200	C
8	18	1800	1	200	C
9	19	2000	1	200	C
10	20	1900	1	-100	D
Total					D=3;C=6

$$f = \frac{C - D}{C + D} = \frac{6 - 3}{6 + 3} = 0{,}3 \; sau \; 30\%$$

Coeficientul de asociere. În cazul variabilelor alternative (dihotomice), datele se sistematizează într-un tabel „2 x 2", care are forma următoare, tabelul 7.18.

Tabelul 7.18

Clasele lui X	**Clasele lui Y**		Total
	$Y(y_1)$	$Y(y_2)$	
0	1	2	3
$X(x_1)$	n_{11}	n_{12}	$n_{1\bullet}$
$X(x_2)$	n_{21}	n_{22}	$n_{2\bullet}$
Total	$n_{\bullet 1}$	$n_{\bullet 2}$	$n_{\bullet\bullet}$

În tabelul 9, $x_1 < \bar{x};\ x_2 > \bar{x};\ \ y_1 < \bar{y};\ y_2 > \bar{y}$ în care, s-a notat cu $\mathbf{x_1}$ toate valorile lui **x** mai mici decât media variabilei aleatoare **X**; cu $\mathbf{x_2}$ toate valorile lui **x** mai mari decât media variabilei aleatoare **X**; cu $\mathbf{y_1}$ toate

Material cu drept de autor

valorile lui **y** mai mici decât media variabilei aleatoare **Y** şi cu y_2 toate valorile lui **y** mai mari decât media variabilei aleatoare **Y**.

O asociere puternică între variabile se remarcă în cazul concentrării frecvențelor pe una din diagonalele tabelului. Dacă toate unitățile statistice sunt dispuse doar pe diagonala principală (n_{11} şi n_{22}), se poate vorbi de o asociere perfect pozitivă (atributul **X** se asociază cu **Y** şi non **X** cu non **Y**), iar dacă unitățile statistice sunt dispuse pe diagonala secundară (n_{21} şi n_{12}), se poate vorbi de o asociere perfect negativă (adică atributul **X** se asociază cu non **Y** şi atributul non **X** se asociază cu atributul **Y**).

Dacă variabilele statistice nu sunt asociate (sunt independente), atunci frecvențele de pe aceeaşi linie şi frecvențele de pe aceeaşi coloană se află în acelaşi raport:

$$\frac{n_{11}}{n_{21}} = \frac{n_{12}}{n_{22}} \Rightarrow n_{11}n_{22} - n_{12}n_{21} = 0$$

Coeficientul φ de măsurare a asocierii dintre variabilele alternative, sistematizate într-un tabel „2 x 2" este:

$$\varphi = \frac{n_{11}n_{22} - n_{21}n_{12}}{\sqrt{n_{\bullet 1}n_{\bullet 2}n_{1\bullet}n_{2\bullet}}} \qquad (7.121)$$

Coeficientul φ ia valori în intervalul **[-1, +1]**. El este o măsură a intensității dependenței dintre clasificarea pe rânduri şi clasificarea pe coloane. O valoare apropiată de 0, ne arată o independență între aceste clasificări. O valoare apropiată de +1 sau de –1, ne arată o dependență între variabile; în acest caz observațiile din rândul 1 tind să fie clasificate în coloana opusă față de observațiile din rândul 2.

Coeficientul Q al lui Yule, care măsoară şi el intensitatea asocierii dintre variabile alternative, are expresia:

$$Q = \frac{n_{11}n_{22} - n_{21}n_{12}}{n_{11}n_{22} + n_{21}n_{12}} \qquad (7.122)$$

Acest indicator ia valori cuprinse între **–1** şi **+1**; el ia valoarea **0** când $n_{11}n_{22} = n_{21}n_{12}$, deci există o independență între variabile; spre deosebire de φ, acest indicator ia valori extreme chiar şi în cazul în care o singură frecvență interioară a tabelului este nulă. O valoare apropiată de **+1** ne arată o asociere pozitivă; iar o valoare apropiată de **–1**, o asociere negativă.

Exemplu de calcul 7.11. Valoarea fondurilor fixe (u.m.) şi numărul de salariați din 10 unități productive se prezintă în coloanele **1** (x_i) şi **2** (y_j).

Se cere să se determine dacă există o legătură între cele două variabile: independentă (x_i) şi dependentă (y_j).

Material cu drept de autor

Rezolvare. Se optează, la început, pentru aplicarea testului lui Yule pentru a observa dacă există legătură între cele două variabile: cauzală (X) şi rezultativă (Y). Se determină valorile medii:

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{1459}{10} = 145{,}9 \; u.m./unit;$$

$$\bar{y} = \frac{\Sigma y_j}{m} = \frac{113}{10} = 11{,}3 \; salariati/unit;$$

Tabelul 7.19

Nr. crt	x_i	y_j	x_i^2	y_j^2	$x_i.y_j$	$\hat{y}_j$	$(y_j - \hat{y}_j)^2$	$(y_j - \bar{y})^2$
0	1	2	3	4	5	6	7	8
1	20	2	400	4	40	3	1	86,43
2	323	21	104329	441	6783	23	4	94,09
3	156	18	24336	324	2808	12	36	44,89
4	180	14	32400	196	2525	14	0	7,29
5	98	11	9604	121	1078	8	9	0,09
6	73	6	5329	36	438	7	1	28,09
7	334	21	111556	441	7014	24	9	94,09
8	20	1	400	1	20	3	4	106,09
9	52	2	2704	4	104	5	9	86,43
10	203	17	41209	289	3451	15	4	32,49
Total	1459	113	332267	1857	24256	113	77	579,98

Se alcătuieşte tabelul de asociere, tabelul 7.20.
Coeficientul lui Yule se determină cu expresia:

$$Q = \frac{n_{11}n_{22} - n_{21}n_{12}}{n_{11}n_{22} + n_{21}n_{12}} = \frac{5.5 - 0}{5.5 + 0} = +1$$

Tabelul 7.20

x_i \ y_i	$y_j < 11{,}3$	$y_j > 11{,}3$	Total
0	1	2	3
$x_i < 145{,}9$	$n_{11} = 5$	$n_{12} = 0$	$n_{1\bullet} = 5$
$x_i > 145{,}9$	$n_{21} = 0$	$n_{22} = 5$	$n_{2\bullet} = 5$
Total	$n_{\bullet 1} = 5$	$n_{\bullet 2} = 5$	$n_{\bullet\bullet} = 10$

Se observă existența unei legături foarte puternice şi pozitive.

Material cu drept de autor

În continuare, pentru a analiza dependența dintre variabila cauzală **X** şi variabila rezultativă **Y**, se aplică un model de regresie liniară simplă, de forma:

$$\hat{y}_j = a + b.x_i$$

Prin aplicarea metodei celor mai mici pătrate rezultă sistemul:

$$\begin{cases} n.a + b.\Sigma x_i = \Sigma y_j \\ a.\Sigma x_i + b.\Sigma x_i^2 = \Sigma x_i y_j \end{cases}$$

de unde se obține:

$$a = \frac{\Sigma y_j \Sigma x_i^2 - \Sigma x_i \Sigma y_j}{n\Sigma x_i^2 - (\Sigma x_i)^2} = \frac{10 \times 332267 - 1459 \times 24256}{10 \times 332267 - (1459)^2} = \frac{2156667}{1193989} = 1{,}80627$$

$$b = \frac{n\Sigma x_i \Sigma y_j - \Sigma x_i \Sigma y_j}{n\Sigma x_i^2 - (\Sigma x_i)^2} = \frac{10 \times 24256 - 1459 \times 113}{10 \times 332267 - (1459)^2} = \frac{77693}{1193989} = 0{,}06507$$

Modelul de regresie va fi:

$$\hat{y}_j = 1{,}80627 + 0{,}06507.x_i$$

Calculele intermediar sunt prezentate în tabelul 7.19. Valorile ajustate ale numărului de salariați în dependență de valoarea fondurilor fixe sunt calculate în coloana 6 din tabelul 7.19.

Cum panta dreptei este b > 0 se trage concluzia că între cele două variabile există o legătură directă. Dacă valoarea fondurilor fixe creşte cu 100 u.m., atunci numărul de salariați creşte în medie cu 6,5 ≈ 7 salariați, dacă nu se implică o automatizare a proceselor de lucru.

Exemplu de calcul 7.12. Înregistrările statistice stabilesc următoarele date privind valoarea fondurilor fixe şi producția globală la opt unități economice cu profil similar.

Tabelul 6.21

Nr. crt	Firma	Valoarea fondurilor fixe (u.m.)	Producția globală (u.m.)
1	F_1	127.296	751.173
2	F_2	342.158	1.123.600
3	F_3	97.925	644.173
4	F_4	286.469	894.013
5	F_5	111.445	803.597
6	F_6	252.483	842.665
7	F_7	124.274	620.277
8	F_8	404.147	964.966

Material cu drept de autor

Se cere să se studieze existența, direcția şi intensitatea legăturii dintre valoare fondurilor fixe (u.m.) şi valoarea producției globale anuale (u.m.), cu ajutorul coeficienților de corelație a rangurilor ai lui Spearman şi Kendall.

Rezolvare. Coeficientul lui Spearman se determină cu expresia:

$$c = 1 - \frac{6\Sigma d_i^2}{n.(n^2 - 1)} \qquad (7.123)$$

unde d_i reprezintă diferența între rangurile celor două variabile, aferente aceleiaşi unități. Se ordonează perechile de valori în ordinea crescătoare a valorilor x_i, apoi se atribuie ranguri valorilor celor două variabile, tabelul 7.22.

Tabelul 7.22

Nr. crt	Firma	Fonduri fixe	Producția globală	R^x	R^y	d_i	d_i^2
3	F_3	97.925	644.173	1	2	-1	1
5	F_5	111.445	803.597	2	4	-2	4
7	F_7	124.274	620.277	3	1	2	4
1	F_1	127.296	751.173	4	3	1	1
6	F_6	252.483	842.665	5	5	0	0
4	F_4	286.469	894.013	6	6	0	0
2	F_2	342.158	1.123.600	7	8	-1	1
8	F_8	404.147	964.966	8	7	1	1
Total							12

$$c = 1 - \frac{6\Sigma d_i^2}{n.(n^2 - 1)} = 1 - \frac{6.12}{8.(8^2 - 1)} = 1 - \frac{6.12}{8.63} = 0{,}86$$

Coeficientul lui Kendall se determină cu expresia:

$$k = \frac{2.(P - Q)}{n.(n - 1)} \qquad (7.124)$$

Se întocmeşte tabelul 7.23:

Tabelul 7.23

Nr. crt	Firma	R^x	R^y	d_i	d_i^2	P	Q	P-Q
3	F_3	1	2	-1	1	6	1	5
5	F_5	2	4	-2	4	4	2	2
7	F_7	3	1	2	4	5	0	5
1	F_1	4	3	1	1	4	0	4
6	F_6	5	5	0	0	3	0	3
4	F_4	6	6	0	0	2	0	2

Material cu drept de autor

2	F_2	7	8	-1	1	0	1	-1
8	F_8	8	7	1	1	0	0	0
Total						24	4	20

Efectuând înlocuirile necesare se obține:

$$k = \frac{2.(P-Q)}{n.(n-1)} = \frac{2.20}{8.7} = 0,71$$

valoare a cărei interpretare este aceeaşi ca în cazul coeficientului lui Spearman.

Exemplu de calcul 7.13. Datele sistematizate, obținute în urma unui studiu privind vechimea în muncă şi timpul zilnic nelucrat, efectuat pe 800 de salariați ai unei societăți comerciale sunt:

Tabelul 7.24

Vechimea	Timp nelucrat	
	Sub 60 u.m.	Peste 60 u.m.
Peste 10 ani	300	150
Sub 10 ani	100	250

Să se precizeze dacă între cele două variabile există o legătură.

Rezolvare. Se foloseşte coeficientul de asociere, calculat cu expresia:

$$Q_a = \frac{a.d - b.c}{a.d + b.c} \qquad (7.125)$$

unde: **a = 300; b = 150; c = 100; d = 250**.

Efectuând înlocuirile necesare rezultă:

$$Q_a = \frac{300.250 - 100.150}{300.250 + 100.150} = 0,67$$

Cum **a.d > b.c** rezultă că între cele două variabile există o legătură de intensitate medie.

Exemplu de calcul 7.14. Pentru două variabile statistice între care există o dependență liniară s-au înregistrat valori experimentale, care în urma prelucrării, se prezintă astfel:

$$\Sigma x_i = 183,91;\ \Sigma y_i = 968;\ \Sigma x_i^2 = 4099,8;$$

$$\Sigma x_i y_i = 1948,9;\ \Sigma y_i^2 = 96299,6;\ i = \overline{1,11}$$

Se cere să se determine:

a. ecuația de regresie;
b. coeficientul de corelație.

Rezolvare:

Estimările dispersiilor variabilelor **X** şi **Y** sunt:

Material cu drept de autor

$$s_x^2 = \frac{\Sigma x_i^2 - \frac{(\Sigma x)^2}{n}}{n-1} = \frac{4099,8 - \frac{(183,91)^2}{11}}{10} = 102,499$$

$$s_y^2 = \frac{\Sigma y_i^2 - \frac{(\Sigma y)^2}{n}}{n-1} = \frac{96299,6 - \frac{(968)^2}{11}}{10} = 1111,56$$

iar abaterile medii pătratice vor fi:

$$s_x = \sqrt{s_x^2} = \sqrt{102,499} = 10,124;\ s_y = \sqrt{s_y^2} = \sqrt{1111,56} = 33,34$$

Covarianța dintre variabilele **X** şi **Y** este:

$$\text{cov}(x,y) = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{n-1} = \frac{19485,9 - \frac{183,91.968}{11}}{10} = 330,182$$

Ca urmare:

$$b = \frac{\text{cov}(x,y)}{s_x^2} = \frac{330,182}{102,499} = 3,2213 \text{ şi}$$

$$a = \bar{y} - b.\bar{x} = \frac{\Sigma y}{n} - b.\frac{\Sigma x}{n} = 88 - 3,2213.16,72 = 34,14$$

Aşadar, **ecuația de regresie este**:

$$\hat{y} = 34,14 + 3,2213.x$$

Coeficientul de corelație $r_{y/x}$ este:

$$r_{y/x} = \frac{\text{cov}(x,y)}{s_x.s_y} \ sau \quad r_{y/x} = b.\frac{s_x}{x_y} \text{ adică:}$$

$$r_{y/x} = \frac{330,182}{10,124.33,34} = 0,978 \ sau$$

$$r_{y/x} = 3,2213.\frac{10,124}{33,34} = 0,978$$

Material cu drept de autor

CAP.8. SERII CRONOLOGICE (SCR)

8.1. Conceptul de serii cronologice

Seria cronologică, numită şi serie dinamică sau serie de timp, este formată dintr-un şir ordonat de valori ale unei variabile, înregistrate pentru momente sau intervale de timp succesive. Se poate simboliza prin [$\mathbf{y_t}$] unde $t = \overline{1,n}$ fiind dezvoltată sub forma:

$$\begin{pmatrix} 1 & 2 & \ldots\ldots & n-1 & n \\ y_1 & y_2 & \ldots\ldots & y_{n-1} & y_n \end{pmatrix}$$

Seriile cronologice se caracterizează prin următoarele particularităţi sau trăsături specifice: **a**. variabilitatea **b**. omogenitatea; **c**. comparabilitatea; **d**. interdependenţa în timp a termenilor.

Variabilitatea termenilor unei serii dinamice apare ca urmare a faptului că fiecare termen se obţine prin centralizarea unor date individuale diferite ca nivel de dezvoltare. Existenţa unor date individuale diferite se explică prin faptul că, în cadrul fenomenelor sociale acţionează, pe lângă cauzele esenţiale, determinante şi un număr suficient de mare de cauze neesenţiale, a căror mod de asociere se poate schimba de la o perioadă la alta.

Omogenitatea presupune că în aceeaşi serie nu pot fi înscrise fenomene de gen diferit, care nu sunt rezultatul acţiunii aceloraşi legi de formare. Condiţia omogenităţii seriilor dinamice nu exclude posibilitatea ca într-un tabel statistic să figureze date care caracterizează perioade calitativ deosebite. Ceea ce este esenţial în elaborarea seriilor dinamice este precizarea exactă a problemei pe care o supunem studiului.

Periodicitatea termenilor presupune alegerea corectă a unităţii de timp la care se referă termenii unei serii cronologice. De exemplu, înregistrările datelor cronologice se fac orar, zilnic, săptămânal, lunar, anual etc.

Interdependenţa termenilor se explică prin aceea că termenii seriei sunt valori succesive ale aceluiaşi fenomen, ca urmare a respectării principiului unităţii de timp şi spaţiu. Aceasta face ca valoarea fiecărui termen să depindă de valoarea termenului anterior ceea ce înseamnă o interdependenţă relativă a termenilor seriei.

8.2. Clasificare seriilor cronologice (SCR)

Există mai multe tipuri de serii cronologice, diferenţiate în funcţie de timpul la care se referă datele, modul de exprimare a indicatorilor, natura fenomenului evidenţiat şi numărul termenilor.

1. În funcţie de modul de definire a timpului, valorile individuale ale seriei cronologice se raportează la un interval sau la un moment de timp.

După acest criteriu deosebim **serii cronologice de intervale** şi **serii cronologice de momente**.

Material cu drept de autor

Seriile cronologice de intervale (de fluxuri) sunt formate din mărimi asociate unor perioade de timp. Fiecare valoare individuală y_i reprezintă rezultatul unui proces care se desfăşoară pe un interval de timp t_i - t_{i+1}, figura 8.1.

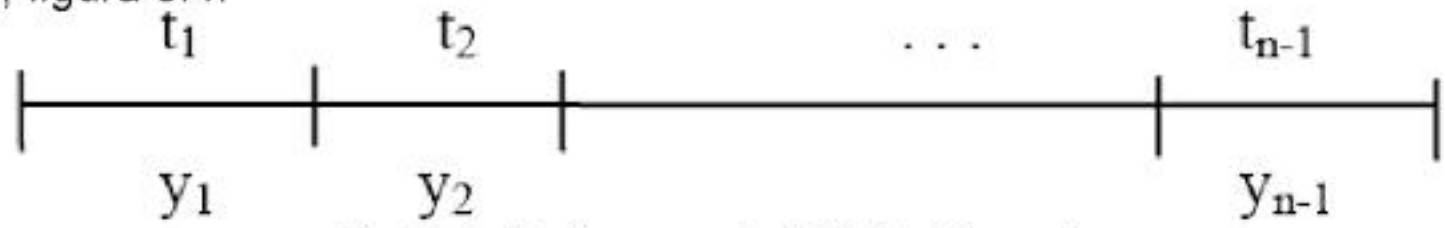


Fig.8.1. Seria cronologică de intervale

Exemple de serii cronologice de intervale: investiţiile anuale realizate de o anumită firmă, cheltuielile lunare de consum ale populaţiei, profiturile trimestriale ale unei societăţi comerciale, desfacerile zilnice de mărfuri ale unei unităţi comerciale sau valoarea tranzacţiilor lunare la bursă.

O proprietate importantă a seriilor cronologice de intervale o reprezintă posibilitatea însumării valorilor $\mathbf{y_i}$; în acest fel se obţine un indicator totalizator pentru întreaga perioadă de timp considerată: $\mathbf{t_1}$ - $\mathbf{t_n}$. De exemplu, prin însumarea desfacerilor zilnice se obţine desfacerea totală lunară, prin cumularea producţiilor lunare se determină producţia anuală etc.

Seriile cronologice de momente (de stocuri) cuprind mărimi care se referă la anumite momente de timp. Fiecare valoare individuală $\mathbf{y_i}$ caracterizează nivelul la care a ajuns fenomenul considerat în momentul de timp $\mathbf{t_i}$, figura 8.2.

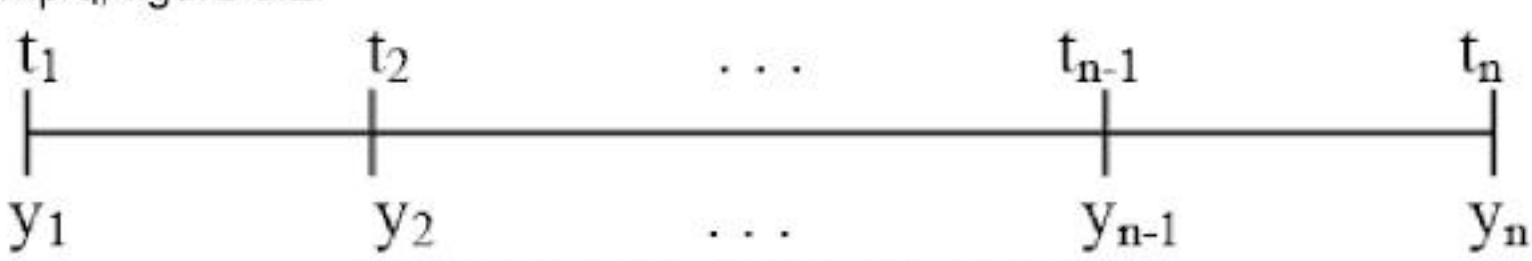


Fig.8.2. Serie cronologică de momente

De exemplu: stocul de materii prime sau de produse finite al unei firme la începutul fiecărei luni, numărul personalului muncitor la sfârşitul fiecărui trimestru, capitalul fix în funcţiune la sfârşitul anului, volumul depozitelor bancare la sfârşitul semestrului etc.

2. În funcţie de modul de exprimare a termenilor seriei deosebim serii cronologice formate din indicatori absoluţi, relativi sau medii.

Seriile cronologice formate din indicatorii absoluţi reprezintă situaţia cea mai frecvent întâlnită. Fiecare termen al seriei este în acest caz o mărime absolută exprimată în unităţi concrete de măsură. De exemplu: producţia zilnică a unei secţii (în unităţi fizice sau valorice), încasările lunare ale unui magazin, valoarea creditelor anuale acordate de o bancă etc.

Seriile cronologice formate din indicatori relativi se exprimă procentual sau sub formă de coeficienţi. Termenii acestor serii reprezintă mărimi relative de structură, de coordonare, de intensitate sau de dinamică. De exemplu: dinamica anuală a PIB (%), ponderea populaţiei ocupate în agricultură (%), cursul zilnic al dolarului (lei/$) sau raportul dobândă activă-

Material cu drept de autor

dobândă pasivă într-o perioadă de timp. Baza de raportare trebuie să fie întotdeauna precizată.

Seriile cronologice formate din indicatori medii se caracterizează prin aceea că termenii seriei sunt calculați ca valori medii. Aceasta este o modalitate de prezentare a evoluției în timp a unor indicatori de moment (transformați în indicatori de intervale prin calcularea nivelului mediu pe fiecare interval între două momente succesive) sau a unor caracteristici calitative. De exemplu: stocurile medii de materiale, mărfuri sau produse finite, productivitatea medie a muncii, salariul mediu, numărul mediu de salariați etc.

3. După numărul termenilor pe care îi conțin, seriile cronologice pot fi de lungime mică, medie sau mare.

Analiza seriilor cronologice urmăreşte frecvent să caracterizeze modul în care a evoluat un fenomen într-o perioadă anterioară, în vederea prognozei evoluției sale probabile în viitor. Atingerea acestui scop presupune parcurgerea câtorva etape:

- constituirea seriei cronologice;
- prelucrarea termenilor seriei cronologice şi obținerea indicatorilor statistici absoluți, relativi şi medii;
- aplicarea metodei ajustării termenilor seriei cronologice în funcție de timp pentru determinarea tendinței pe termen lung;
- determinarea influenței factorilor sezonieri şi analiza fenomenelor cu caracter ciclic;
- estimarea valorilor probabile pentru perioada următoare (prin metoda extrapolării).

Desfăşurarea acestor etape de calcul şi analiză prezintă particularități în funcție de tipul seriei cronologice.

8.3. Analiza seriilor cronologice de intervale

Întrucât termenii seriei cronologice prezintă variații mari de la o perioadă de timp la alta, prima fază, obligatorie în studiul oricărei serii, o reprezintă calcularea unui sistem de indicatori statistici absoluți, relativi şi medii. Aceşti indicatori caracterizează modificarea în timp a fenomenului analizat.

Indicatori absoluți. Indicatorii absoluți ai unei serii cronologice de intervale exprimă nivelul, volumul agregat şi modificările (în mărime absolută) fenomenului analizat în perioade diferite de timp. Indicatorii absoluți se exprimă în unitatea de măsură a caracteristicii analizate (în unități fizice, valorice, procente etc.).

Valorile individuale absolute ale caracteristicii redau nivelul y_t al fenomenului analizat în fiecare interval de timp t_i (vezi coloana 1 a tabelului 8.1).

Volumul agregat (nivelul totalizat) reprezintă suma termenilor seriei cronologice de intervale: Σy_t**=2146**, coloana 1, tabelul 8.1

Material cu drept de autor

Modificarea absolută (sporul sau scăderea absolută) reflectă creşterea sau descreşterea absolută (în unități concrete de măsură) a valorilor individuale ale fenomenului analizat, de la o perioadă de timp la alta.

Modificarea absolută se calculează ca diferență între doi termeni ai seriei. În funcție de perioada aleasă ca bază de comparație (constantă sau variabilă), există două forme ale acestui indicator:

• modificarea absolută cu bază fixă reprezintă distanța (diferența) fiecărui termen al seriei față de o perioadă fixă de referință:

Diferențele absolute se simbolizează prin $\Delta^{y}_{k/j}$ şi se calculează scăzând din volumul atins de fenomenul **y** în momentul sau pe intervalul **k**, volumul aceluiaşi fenomen înregistrat în momentul sau pe intervalul **j** luat ca bază de comparație:

$$\Delta^{y}_{k/j} = y_k - y_j \qquad (8.1)$$

Valoarea numerică a unui astfel de indicator arată cu cât a crescut sau a scăzut volumul fenomenului cercetat **y** în intervalul sau la momentul **k** față de intervalul sau momentul **j**.

- diferențele absolute cu bază fixă se obțin prin compararea fiecărui termen al seriei cu unul şi acelaşi termen – care poate fi primul sau oricare altul ales în urma unei analize atente în funcție de scopul urmărit. Dacă **j = 1**:

$$\Delta^{y}_{k/j} = y_k - y_1 \quad (k = 2,3,\ldots,n) \qquad (8.2)$$

- diferențele absolute cu bază în lanț (mobilă) se obțin prin compararea fiecărui termen al seriei cu precedentul său:

$$\Delta^{y}_{k/k-1} = y_k - y_{k-1} \qquad (k = 2,3,\ldots,n;\ j = k-1) \qquad (8.3)$$

Aceşti indicatori, spre deosebire de diferențele cu bază fixă, arată cu cât a crescut sau s-a micşorat volumul fenomenului **y** de la un moment la următorul (o lună, trimestru, an etc).

Pe baza relațiilor (8.2) şi (8.3), se pot forma expresii de calcul pentru trecerea de la diferențele cu bază în lanț la diferențele cu bază fixă şi invers. Astfel:

- prin cumularea diferențelor absolute cu baza în lanț se obțin diferențele absolute cu baza fixă:

$$\sum_{k=2}^{m} \Delta^{y}_{k/k-1} = \Delta^{y}_{m/1} \quad (m = 2,3,\ldots,n) \qquad (8.4)$$

De exemplu:

$$\Delta^{y}_{2/1} + \Delta^{y}_{3/2} = (y_2 - y_1) + (y_3 - y_2) = (y_3 - y_1) = \Delta^{y}_{3/1}$$

$$\Delta^{y}_{2/1} + \Delta^{y}_{3/2} + \Delta^{y}_{4/3} = \Delta^{y}_{3/1} + \Delta^{y}_{4/3} = (y_2 - y_1) + (y_3 - y_2) + (y_4 - y_3) = \Delta^{y}_{4/1}$$

- prin scăderi succesive ale diferențelor absolute cu baza fixă se obțin diferențele absolute cu baza în lanț:

Material cu drept de autor

$$\Delta^y_{k/1} - \Delta^y_{k-1/1} = \Delta^y_{k/k-1} \qquad (8.5)$$

De exemplu:

$$\Delta^y_{3/1} - \Delta^y_{2/1} = (y_3 - y_1) - (y_2 - y_1) = (y_3 - y_2) = \Delta^y_{3/2}$$

Modul de calcul al diferențelor cu bază fixă şi cu bază în lanț este prezentat în tabelul 8.1, coloanele 2 şi 3, iar pentru interpretare se poate considera:

$$\Delta^y_{2/1} = y_2 - y_1 = 192 - 135 = 57 \text{ mii buc.}$$

$$\Delta^y_{6/1} = y_6 - y_1 = 332 - 135 = 197 \text{ mii buc.}$$

Față de primul an (1), producția produsului (**y**) a crescut în anul (2) cu 57 mii buc., iar în anul (6) cu 197 mii buc.

Producția produsului (**y**) a crescut în anul (7) față de anul precedent cu 53 mii buc.

$$\Delta^y_{7/6} = y_7 - y_6 = 332 - 279 = 53 \text{ mii buc.}$$

Acelaşi rezultat se obține şi prin scăderea diferențelor cu bază în lanț ale celor doi ani:

$$\Delta^y_{7/6} = \Delta^y_{7/1} - \Delta^y_{6/1} = 197 - 144 = 53 \text{ mii buc.}$$

Calculați în coeficienți sau în procente, indicii arată de câte ori sau în ce proporția s-a modificat fenomenul **y** pe perioada considerată.

Dacă baza de comparație (**j**) se menține constantă sau se schimbă, ca şi diferențele absolute, indicii pot fi calculați cu baza fixă sau cu baza în lanț.

Indici relativi. Aceşti indicatori se calculează sub formă de raport şi reflectă proporția dintre nivelurile absolute ale termenilor seriei cronologice de intervale. Permit analiza comparativă a evoluției unor fenomene diferite. Ca urmare **indicele de dinamică** este o mărime relativă care arată de câte ori s-a modificat mărimea unui fenomen în timp. Se calculează ca raport între doi termeni diferiți ai seriei cronologice.

Indicii simbolizați prin $I^y_{k/j}$ se obțin prin raportarea volumului înregistrat de fenomenul **y** în momentul sau intervalul **k** la volumul aceluiaşi fenomen din momentul sau intervalul **j** luat ca bază de comparație:

$$I^y_{k/j} = \frac{y_k}{y_j} \text{ sau } I^y_{k/j} = \frac{y_k}{y_j} 100 \qquad (8.6)$$

În funcție de alegerea unei baze de raportare constante sau variabile, se poate determina indicele cu bază fixă sau cu bază în lanț.

Indicii cu bază fixă se obțin prin raportarea fiecărui termen al seriei la aceeaşi bază de comparație (de obicei se alege ca bază primul termen al seriei):

$$I^y_{k/1} = \frac{y_k}{y_1} \quad (k = 1,2,...,n;\ j = 1) \qquad (8.7)$$

Material cu drept de autor

Indicii cu bază în lanț se obțin prin raportarea fiecărui termen al seriei la precedentul său:

$$I^{y}_{k/k-1} = \frac{y_k}{y_{k-1}} \quad (k = 2,3,\ldots,n;\ j = k-1) \tag{8.8}$$

Între indicii cu bază fixă şi indicii cu baza în lanț există relații care permit calculul unora din ceilalți. Astfel:

- prin înmulțirea indicilor cu baza în lanț, până la un anumit moment sau interval, se obține indicele cu bază fixă al acelui moment sau interval:

$$\prod_{k=2}^{m} I^{y}_{k/k-1} = I^{y}_{m/1} \quad (m = 2,3,\ldots,m) \tag{8.9}$$

De exemplu:

$$I^{y}_{2/1}.I^{y}_{3/2} = \frac{y_2}{y_1} \times \frac{y_3}{y_2} = \frac{y_3}{y_1} = I^{y}_{3/1}$$

- prin împărțiri succesive ale indicilor cu bază fixă se obțin indici cu baza în lanț:

$$I^{y}_{k/1} : I^{y}_{k-1/1} = I^{y}_{k/k-1} \quad (k = 2,3,\ldots,n) \tag{8.10}$$

De exemplu:

$$I^{y}_{3/1} : I^{y}_{2/1} = \frac{y_3}{y_1} : \frac{y_2}{y_1} = I^{y}_{3/2}$$

Indicii cu bază fixă şi în lanț calculați pentru a caracteriza dinamica producției produsului (**y**) pe perioada de opt ani sunt prezentați în coloanele 4 şi 5 din tabelul 8.1.

$$I^{y}_{2/1} = \frac{y_2}{y_1}100 = 142,2\%$$

$$I^{y}_{6/1} = \frac{y_6}{y_1}100 = \frac{332}{135}100 = 245,9\%$$

Față de primul an, producția a crescut în al doilea an în proporție de 142,2% (sau de 1,422 de ori), iar în anul şase în proporție de 245,9% (sau de 2,459 ori).

$$I^{y}_{6/5} = \frac{y_6}{y_5}100 = \frac{332}{279}100 = 119,0\%$$

sau:

$$I^{y}_{6/5} = I^{y}_{6/1} : I^{y}_{5/1} = 2,459 : 2,067 = 1,19$$

Material cu drept de autor

Tabelul 8.1

Evoluția şi dinamica producției produsului (y) pe o perioadă de opt ani

Anii (t_k)	Producția (mii buc.), y_k	Diferențe absolute		Indicii, %		Ritmurile, %		Valoarea absolută a unui procent din ritm	
		Cu bază fixă $\Delta^y_{k/1}$	Cu bază în lanț $\Delta^y_{k/k-1}$	Cu bază fixă $I^y_{k/1}$	Cu bază în lanț $I^y_{k/k-1}$	Cu bază fixă $R^y_{k/1}$	Cu bază în lanț $R^y_{k/k-1}$	Cu bază fixă $A^y_{k/1}$	Cu bază în lanț $A^y_{k/k-1}$
0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	135	-	-	100,0	-	-	-	-	-
2	192	57	57	142,2	142,2	42,2	42,2	1,35	1,35
3	196	61	4	145,2	102,1	45,2	2,1	1,35	1,92
4	222	87	26	164,4	113,3	64,4	13,3	1,35	1,96
5	279	144	57	206,7	125,7	106,7	25,7	1,35	2,22
6	332	197	53	245,9	119,0	145,9	19,0	1,35	2,79
7	376	241	44	278,5	113,3	178,5	13,3	1,35	3,32
8	414	279	38	306,7	110,1	206,7	10,1	1,35	3,76
	$\sum_{k=1}^{8} y_k$ = 2146	X	$\sum_{k=2}^{8} \Delta^y_{k/k-1}$ = 279	X	$\prod_{k=2}^{8} I^y_{k/k-1}$ = 3,067	X	X	X	X

Material cu drept de autor

Ritmul de dinamică (de creştere sau scădere), numit şi ritmul modificării arată cu cât s-a modificat procentual (a crescut sau a scăzut) mărimea fenomenului într-o anumită perioadă de timp față de o perioadă de referință fixă sau mobilă. Se determină ca raport între modificarea absolută (cu bază fixă sau în lanț) şi nivelul fenomenului în perioada aleasă ca termen de comparație.

Simbolizați prin $R^{y}_{k/j}$, aceşti indicatori se obțin prin raportarea diferenței absolute (sporului absolut) la nivelul atins de fenomenul **y** în momentul sau pe intervalul **j** luat ca bază de comparație:

$$R^{y}_{k/j} = \frac{\Delta^{y}_{k/j}}{y_j} \quad \text{sau} \quad R^{y}_{k/j} = \frac{\Delta^{y}_{k/j}}{y_j}100 \tag{8.11}$$

Prin desfăşurarea numărătorului se ajunge la calculul ritmurilor în funcție de indici:

$$R^{y}_{k/j} = \frac{\Delta^{y}_{k/j}}{y_j} = \frac{y_k - y_j}{y_j} = \frac{y_k}{y_j} - 1 = I^{y}_{k/j} - 1$$

Aşadar:

$$R^{y}_{k/j} = \frac{\Delta^{y}_{k/j}}{y_j}.100 \quad \text{sau} \quad R^{y}_{k/j} = I^{y}_{k/j} - 100 \tag{8.12}$$

Ritmurile ca şi indicii se pot exprima în coeficienți sau în procente şi arată cu cât în cifre relative (față de 1 sau 100) s-a modificat volumul fenomenului **y** în momentul sau pe intervalul **k** față de momentul sau intervalul **j** bază de comparație.

În funcție de scopul urmărit în cercetare, baza de comparație **j** se poate menține constantă sau se schimbă şi astfel pot fi calculate ritmurile cu bază fixă şi ritmurile cu bază în lanț.

Ritmurile cu bază fixă:

$$R^{y}_{k/1} = \frac{\Delta^{y}_{k/1}}{y_1}.100 \quad \text{sau} \quad R^{y}_{k/1} = I^{y}_{k/1}\% - 100 \tag{8.13}$$

k = 2, 3,..., n

oglindesc creşterea sau descreşterea relativă a fenomenului în fiecare moment următor față de acelaşi moment luat ca bază de comparație.

Ritmurile cu bază în lanț:

$$R^{y}_{k/k-1} = \frac{\Delta^{y}_{k/k-1}}{y_{k-1}} \quad \text{sau} \quad R^{y}_{k/k-1} = I^{y}_{k/k-1}\% - 100 \tag{8.14}$$

k = 2, 3,..., n; j = k – 1

cuantifică creşterile sau descreşterile relative ale fenomenului **y** de la un an la altul sau de la un moment la altul.

Material cu drept de autor

Calculul ritmurilor cu bază fixă şi în lanț pentru a aprecia creşterile relative ale producției produsului (**y**) pe o perioadă de opt ani este prezentat în tabelul 8.1, coloanele 6 şi 7, iar pentru interpretare concretă exemplificăm prin efectuare calculelor:

$$R^{y}_{6/1} = \frac{\Delta^{y}_{6/1}}{y_1}.100 = \frac{197}{135}100 = 145{,}9\%$$

sau:

$$R^{y}_{6/1} = I^{y}_{6/1}\% - 100 = 245{,}9 - 100 = 145{,}9\%$$

respectiv:

$$R^{y}_{6/5} = \frac{\Delta^{y}_{6/5}}{y_5}100 = \frac{53}{279}.100 = 19{,}0\%$$

sau

$$R^{y}_{6/5} = I^{y}_{6/5}\% - 100 = 119{,}0 - 100 = 19{,}0\%$$

ceea ce înseamnă că producția în anul şase a crescut cu 145,9% față de primul an şi cu 19,0% față de anul precedent, cinci.

Ritmurile ca şi indicii se pot utiliza nu numai pentru caracterizarea dinamicii fenomenelor social-economice pe diferite perioade de timp, ci şi pentru stabilirea unor posibilități de dezvoltare a unor indicatori şi pentru aprecierea gradului de realizare a uni program prestabilit.

Valoarea absolută a unui procent de creştere sau de scădere ($A^{y}_{k/j}$) arată mărimea absolută a modificării ce revine pe o unitate (un procent) din ritmul dinamicii. Se calculează sub forma unui raport între modificarea absolută şi ritmul modificării şi se exprimă în unitatea de măsură a caracteristicii.

Acest indicator face legătura dintre indicatorii absoluți şi cei relativi:

$$A^{y}_{k/j} = \frac{\Delta^{y}_{k/j}}{R^{y}_{k/j}.100} = \frac{y_k - y_j}{\frac{y_k - y_j}{y_j}.100} = \frac{y_j}{100} \qquad (8.15)$$

Ca şi ceilalți indicatori şi aceştia se pot calcula cu bază fixă sau cu bază în lanț (mobilă).

Valoarea absolută a unui procent din ritmul cu bază fixă:

$$A^{y}_{k/1} = \frac{\Delta^{y}_{k/1}}{R^{y}_{k/1}.100} = \frac{y_1}{100} \qquad (8.16)$$

are un conținut de medie fiind constant pe întreaga perioadă.

În exemplul considerat în tabelul 8.1, coloanele 8 şi 9:

$$A^{y}_{k/1} = \frac{135}{100} = 1{,}35$$

Material cu drept de autor

arată că la creşterea cu 1% a producției în fiecare interval față de primul interval revine o creştere absolută de 1,35 bucăți.

Valoarea absolută a unui procent din ritmul cu baza în lanț:

$$A^{y}_{k/k-1} = \frac{\Delta^{y}_{k/k-1}}{R^{y}_{k/k-1}.100} = \frac{y_{k-1}}{100} \quad (k = 2,3,...,n) \tag{8.17}$$

nu mai are caracter de medie, ci diferă de la un interval la altul.

În exemplul de mai sus:

$$A^{y}_{6/5} = \frac{y_5}{100} = \frac{279}{100} = 2,79 \approx 2,8 \text{ buc.}$$

ceea ce înseamnă că la o creştere cu 1% a producției în al şaselea interval față de cel precedent îi revine o creştere în cifre absolute de 2,8 buc.

Determinarea nivelului mediu al seriei. Un indicator important ce caracterizează sintetic nivelul seriei dinamice este nivelul mediu al seriei cronologice.

Spre deosebire de indicatorii absoluți şi relativi, care se determină sub forma unui şir de valori care arată variabilitatea termenilor seriei cronologice de intervale, indicatorii medii oferă o măsură sintetică a tendinței de evoluție a întregii serii.

Se pot calcula atât medii de *nivel,* cum sunt media termenilor seriei şi media modificărilor absolute, cât şi medii de *dinamică*: indicele mediu şi ritmul mediu.

Nivelul mediu al termenilor seriei cronologice de intervale se calculează sub forma mediei aritmetice simple a tuturor termenilor seriei cronologice analizate:

$$\bar{y} = \frac{\sum_{t=1}^{n} y_t}{n} = \frac{y_1 + y_2 + ... + y_n}{n} \tag{8.18}$$

unde: **n** - numărul termenilor seriei.

Modificarea medie absolută este media aritmetică simplă a modificărilor absolute cu bază în lanț:

$$\bar{\Delta} = \frac{\sum_{T=2}^{N} \Delta_{t/t-1}}{n-1} \tag{8.19}$$

unde: **n-1** - numărul modificărilor absolute cu bază în lanț.

Întrucât:

$$\sum_{t=2}^{n} \Delta_{t/t-1} = \Delta_{n/1}$$

rezultă o formulă directă de calcul a modificării medii absolute:

$$\bar{\Delta} = \frac{y_n - y_1}{n-1} \tag{8.20}$$

Material cu drept de autor

Indicatorul arată cu cât creşte/scade fenomenul în medie (în valoare absolută) de la o perioadă de timp la alta.

Calcularea acestui indicator are sens atunci când modificările absolute cu bază în lanț nu diferă prea mult ca mărime. **Modificarea medie absolută poartă numele de spor mediu, dacă este calculată pentru un fenomen cu tendință de creştere. În caz contrar vorbim despre scădere medie**.

Indicele mediu de dinamică se calculează ca medie geometrică simplă a indicilor de dinamică cu bază în lanț, conform relației:

$$\bar{I} = \sqrt[n-1]{\prod_{t=2}^{n} I_{t/t-1}} \qquad (8.21)$$

unde: **n-1** - numărul indicilor cu bază în lanț

Întrucât:

$$\prod_{t=2}^{n} I_{t/t-1} = I_{n/1}$$

rezultă o expresie echivalentă:

$$\bar{I} = \sqrt[n-1]{\frac{y_n}{y_1}} \qquad (8.22)$$

Indicele mediu de dinamică arată de câte ori s-a modificat (a crescut sau a scăzut) în medie fenomenul analizat pe întreaga perioadă luată în calcul. Valoarea rezultată din calcul este semnificativă îndeosebi pentru fenomenele care evoluează în progresie geometrică (indicii cu bază în lanț au valori apropiate între ele).

Valori mai mari de 100 % ale acestui indicator arată tendința de creştere a fenomenului analizat. Valori mai mici de 100 % corespund unei scăderi pe ansamblul perioadei considerate.

Atunci când se cunosc indicii medii de dinamică ($\bar{I}_i$) calculați pentru intervalele succesive de timp ***i*** care compun perioada analizată, indicele mediu general se poate calcula direct din aceştia. În acest caz se foloseşte formula mediei geometrice ponderate:

$$\bar{\bar{I}} = \sqrt[\sum_{i=1}^{k} n_i]{\bar{I}_1^{n_1} . \bar{I}_2^{n_2} \ldots \bar{I}_k^{n_k}} \qquad (8.23)$$

unde: $\bar{\bar{I}}$ - indicele mediu de dinamică pe ansamblul perioadei analizate: $\bar{I}_i$ - indicii medii parțiali de dinamică; n_i – numărul indicilor cu bază în lanț care compun indicele mediu $\bar{I}_i$; k – numărul indicilor medii parțiali.

Ritmul mediu de dinamică se determină prin intermediul relației:

$$\bar{R} = \bar{I}\% - 100 \qquad (8.24)$$

Material cu drept de autor

şi arată cu câte procente se modifică în medie fenomenul analizat pe întreaga perioadă analizată.

Aceşti indicatori impun exigenţele comune tuturor indicatorilor medii. Ei nu sunt reprezentativi pentru seriile cronologice cu un număr mare de termeni şi cu oscilaţii mari. În cazul seriilor cronologice cu un nivel scăzut de omogenitate se recomandă separarea seriei pe subperioade care pot fi analizate independent.

8.4. Analiza seriilor cronologice de momente

Spre deosebire de seriile cronologice de intervale, termenii seriilor cronologice de momente se referă la un moment fix, nu la un interval de timp. Distanţele care separă aceste momente de timp pot avea mărime constantă sau variabilă. După acest criteriu clasificăm seriile cronologice de momente în serii cu intervale egale şi cu intervale inegale.

1. Seriile cronologice cu intervale egale între momente pot fi prelucrate în mod asemănător seriilor cronologice de intervale. Se pot calcula indicatorii absoluţi, relativi şi medii prezentaţi anterior.

Excepţie face calculul mediei aritmetice a termenilor seriei. Întrucât termenii seriilor cronologice de momente nu sunt direct însumabili, media se calculează după o formulă specială, ca **medie cronologică simplă**. Pentru fiecare interval dintre două momente succesive se calculează o medie aritmetică simplă a termenilor care mărginesc intervalul, figura 8.3.

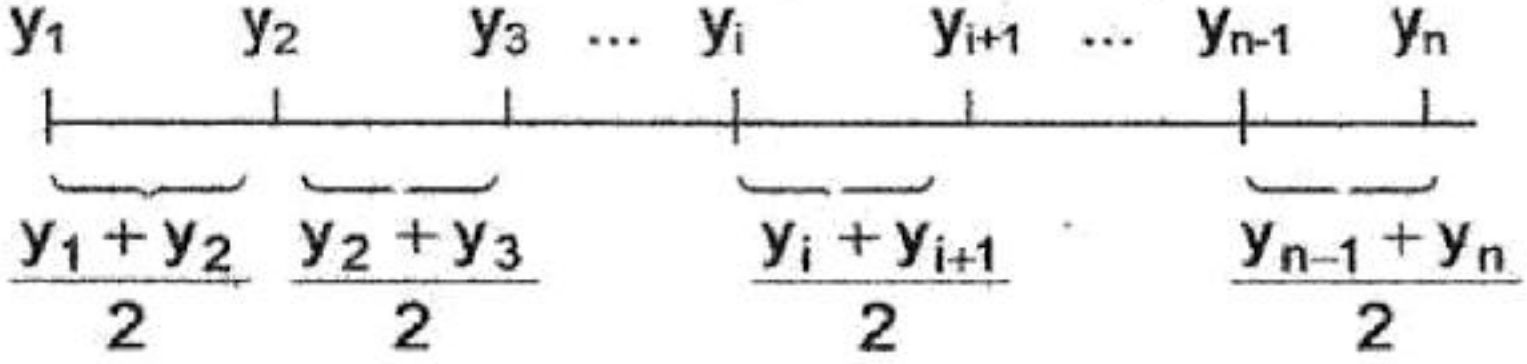


Fig.8.3. Medii parţiale pe intervale egale

Pe ansamblul seriei se calculează o medie aritmetică simplă a acestor ***n-1*** medii parţiale, conform relaţiei:

$$\bar{y}_{cr} = \frac{\frac{y_1 + y_2}{2} + \frac{y_2 + y_3}{2} + \ldots + \frac{y_{n-1} + y_n}{2}}{n-1} = \tag{8.25}$$

$$= \frac{\frac{y_1}{2} + y_2 + y_3 + \ldots + y_{n-1} + \frac{y_n}{2}}{n-1}$$

Pentru prelucrarea seriilor cronologice de momente cu intervale egale între momente se mai poate recurge şi la procedeul transformării acestora în serii de intervale prin *calcularea unei medii aritmetice simple*

Material cu drept de autor

pentru fiecare interval în parte. Seria astfel obținută poate fi prelucrată aşa cum s-a arătat anterior, la seriile cronologice de intervale.

2. Pentru **seriile cronologice cu intervale inegale între momente** este posibilă calcularea unui singur indicator mediu: nivelul mediu al termenilor seriei.

Acest calcul se efectuează după o formulă specială: **media cronologică ponderată.** Întrucât distanțele ce separă momentele de timp la care se referă valorile absolute ale seriei sunt inegale, mediile parțiale pe intervale se ponderează cu mărimea intervalelor respective, figura 8.4.

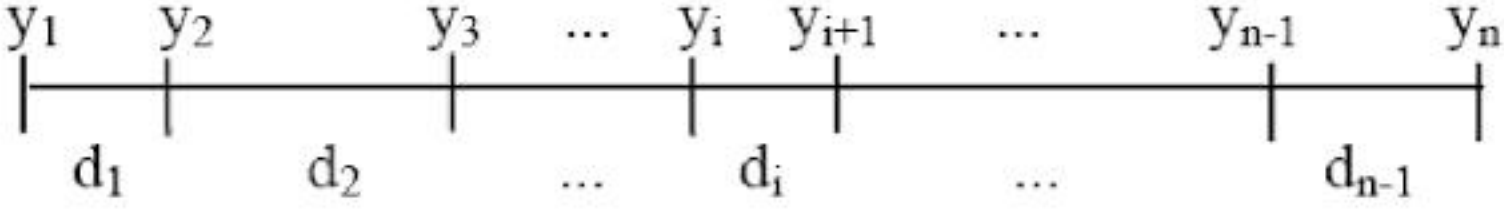


Fig.8.4. Serii cronologice cu intervale inegale între momente

Pe ansamblul perioadei luate în calcul se determină o medie aritmetică ponderată a mediilor de interval:

$$\bar{y}_{cr} = \frac{\frac{y_1+y_2}{2}\times d_1 + \frac{y_2+y_3}{2}\times d_2 + \ldots + \frac{y_{n-2}+y_{n-1}}{2}\times d_{n-2} + \frac{y_{n-1}+y_n}{2}\times d_{n-1}}{d_1+d_2+\ldots+d_{n-1}}$$

$$\Rightarrow \bar{y}_{cr} = \frac{y_1\times\frac{d_1}{2} + y_2\times\frac{d_1+d_2}{2} + \ldots + y_{n-1}\times\frac{d_{n-2}+d_{n-1}}{2} + y_n\frac{d_{n-1}}{2}}{d_1+d_2+\ldots+d_{n-1}} \quad (8.26)$$

unde: $d_1, d_2, \ldots, d_{n-1}$ - mărimea intervalelor dintre momentele de timp la care se referă termenii $y_1, y_2, \ldots, y_n$.

După cum se poate observa în formula finală, valorile individuale ale termenilor se ponderează cu câte o jumătate din mărimea celor două intervale alăturate. Excepție fac termenii extremi, pentru care există un singur interval alăturat.

Datele nefiind comparabile din punctul de vedere al variației în timp, media cronologică ponderată este singurul indicator care se poate calcula în cazul seriilor cronologice de momente inegal distanțate.

Exemplu de calcul 8.1. Pentru primele şase luni ale anului **X**, stocul unui grup omogen de mărfuri a înregistrat următoarea evoluție, tabelul 8.2:

Tabelul 8.2. Date inițiale

Momentul înregistrării	1.I	15.II	1.III	1.IV	1.VI
Valoarea stocului, u.m.	1200	1400	1300	1100	1050

Se cere:

1. Să se precizeze felul seriei;
2. Să se calculeze nivelul mediu al seriei.

Material cu drept de autor

(Pentru simplificarea calculelor se va considera luna egală cu 30 de zile).

Rezolvare.

1. Este o serie de momente cu intervale neegale.
2. Nivelul mediu al seriei ($\bar{y}$) se va calcula cu relația:

$$\bar{y}_{cr} = \frac{y_1 \frac{d_1}{2} + y_2 \frac{d_1 + d_2}{2} + \ldots + y_n \frac{d_{n-1}}{2}}{\frac{d_1}{2} + \frac{d_1 + d_2}{2} + \ldots + \frac{d_{n-1}}{2}}$$

Se vor stabili distanțele dintre momentele seriei, respectiv valorile timpului: $\mathbf{d_1 = 45;\ d_2 = 45;\ d_3 = 30;\ d_4 = 60}$ zile.

Deci:

$$\bar{y}_{cr} = \frac{y_1 \frac{d_1}{2} + y_2 \frac{d_1 + d_2}{2} + y_3 \frac{d_2 + d_3}{2} + y_4 \frac{d_3 + d_4}{2} + y_4 \frac{d_4}{2}}{\frac{d_1}{2} + \frac{d_1 + d_2}{2} + \frac{d_2 + d_3}{2} + \frac{d_3 + d_4}{2} + \frac{d_{n-1}}{2}} =$$

$$= \frac{1200 \frac{45}{2} + 1400 \frac{45 + 45}{2} + 1300 \frac{45 + 30}{2} + 1100 \frac{30 + 60}{2} + 1050 \frac{60}{3}}{\frac{45}{2} + \frac{45 + 45}{2} + \frac{45 + 30}{2} + \frac{30 + 60}{2} + \frac{60}{3}} =$$

$$= \frac{219750}{180} = 1220{,}8\ u.m.$$

8. 5. Ajustarea seriilor cronologice

Specifică seriilor cronologice este variabilitatea mare a termenilor. Aceste variații sunt produse de factori esențiali şi întâmplători. Acțiunea factorilor esențiali determină tendința majoră de evoluție în timp a mărimilor înregistrate de fenomenul analizat. Această tendință (***trend***[1]) interferează cu cauzele neesențiale, întâmplătoare, efectul obținut fiind valorile reale ale fenomenului. Scopul ajustării seriilor cronologice îl reprezintă evidențierea factorilor esențiali, cu acțiune sistematică, care urmăresc o legitate matematică de evoluție.

Ajustarea este operația de înlocuire a termenilor reali ai seriei cronologice cu termeni teoretici care exprimă o anumită legitate matematică de evoluție a fenomenului considerat. Pentru seriile cronologice, această legitate de evoluție se realizează în funcție de timp.

Întrucât abaterea termenilor reali de la cei teoretici calculați este efectul cauzelor neesențiale, întâmplătoare, prin ajustare se evidențiază mai bine tendința de evoluție în timp a fenomenului.

---

[1] În terminologia statistică internațională a căpătat circulație expresia ***trend*** din limba engleză, mai ales în țările de limbă engleză şi germană. Unii autori folosesc şi expresiile ***tendință seculară*** sau ***tendință de lungă durată***, deşi sunt departe de a oglindi noțiunea ce o exprimă.

Material cu drept de autor

Există mai multe procedee de ajustare:

- ajustarea prin metoda mediilor mobile;
- ajustarea prin metoda grafică;
- ajustarea prin metoda modificării mediei absolute;
- ajustarea prin metoda indicelui mediu de dinamică;
- ajustarea prin metode analitice.

Primele patru procedee formează grupul metodelor mecanice de ajustare.

- **Ajustarea pe baza mediilor mobile**. Atunci când graficul seriei cronologice relevă oscilații periodice de la tendința centrală (grafic sinusoidal), este indicată ajustarea pe baza calculului mediilor mobile.

Mediile mobile sunt medii aritmetice parțiale, alunecătoare, calculate din doi sau mai mulți termeni succesivi ai seriei. Se calculează înlocuind pe rând primul termen cu termenul următor din seria cronologică.

Numărul termenilor din care se calculează mediile parțiale se alege în funcție de periodicitatea oscilațiilor seriei. Aceasta poate fi evidențiată cu ajutorul reprezentării grafice, observând mărimea distanței medii dintre punctele de inflexiune ale graficului. În general, atunci când se dispune de date lunare mediile parțiale se calculează din câte 12 termeni succesivi, iar atunci când se folosesc date trimestriale mediile parțiale se calculează din câte 4 termeni succesivi.

Mediile mobile asigură compensarea abaterilor, a oscilațiilor periodice. Noua serie obținută prin ajustare are o variație lină, continuă, evidențiind tendința de evoluție a fenomenului (**trendul**), independent de acțiunea factorilor sezonieri.

Pentru **mediile mobile calculate dintr-un număr impar de termeni** (tabelul 8.3), formula de calcul utilizată este:

$$\bar{y}_i = \frac{y_i + y_{i+1} + \ldots + y_{i+(k-1)}}{k} \qquad (8.27)$$

Tabelul 8.3

Calcule intermediare pentru ajustarea seriilor cronologice prin metoda mediilor mobile dintr-un număr impar de termeni

Valorile absolute (y_i)	Mediile mobile ($\bar{y}_i$)
y_1	-
y_2	$\bar{y}_1 = (y_1 + y_2 + y_3) : 3$
y_3	$\bar{y}_2 = (y_2 + y_3 + y_4) : 3$
M	M
y_i	$\bar{y}_{i-1} = (y_{i-1} + y_i + y_{i+1}) : 3$
M	M

Material cu drept de autor

y_{n-1}	$\bar{y}_{n-1} = (y_{n-2} + y_{n-1} + y_n) : 3$
y_n	-

Numărul mediilor mobile de **k** termeni care se pot calcula dintr-o serie de lungime **n** este: **n - (k - 1)**. Aşadar, în urma ajustării pe baza mediilor mobile se pierd k - 1 termeni ai seriei (la începutul şi la sfârşitul şirului); acesta este principalul dezavantaj al metodei.

În cazul **mediilor mobile calculate dintr-un număr par de termeni** (tabelul 8.4), calculul se realizează în două faze:

1. se obțin medii mobile provizorii din câte k termeni succesivi (folosind formula anterioară; k este număr par în acest caz) care se plasează între termenii seriei reale.

Tabelul 8.4

Modelul de calcul pentru mediile mobile dintr-un număr par de termeni k (k = 4)

Valorile absolute (y_i)	Mediile mobile provizorii ($\bar{y}_i$)	Mediile mobile definitive ($\bar{\bar{y}}_i$)
y_1	–	–
y_2	$\bar{y}_1 = (y_1 + y_2 + y_3 + y_4) : 4$	–
y_3	$\bar{y}_2 = (y_2 + y_3 + y_4 + y_5) : 4$	$\bar{\bar{y}}_1 = (\bar{y}_1 + \bar{y}_2) : 2$
y_4	$\bar{y}_3 = (y_3 + y_4 + y_5 + y_6) : 4$	$\bar{\bar{y}}_2 = (\bar{y}_2 + \bar{y}_3) : 2$
M	M	M
y_{n-3}	$\bar{y}_{n-5} = (y_{n-5} + y_{n-4} + y_{n-3} + y_{n-2}) : 4$	$\bar{\bar{y}}_{n-5} = (\bar{y}_{n-5} + \bar{y}_{n-4}) : 2$
y_{n-2}	$\bar{y}_{n-4} = (y_{n-4} + y_{n-3} + y_{n-2} + y_{n-1}) : 4$	$\bar{\bar{y}}_{n-4} = (\bar{y}_{n-4} + \bar{y}_{n-3}) : 2$
y_{n-1}	$\bar{y}_{n-3} = (y_{n-3} + y_{n-2} + y_{n-1} + y_n) : 4$	–
y_n	–	–

2. se determină medii mobile definitive (centrate) din câte două medii mobile provizorii succesive care se plasează în dreptul termenilor reali (pe care îi înlocuiesc).

Numărul termenilor reali care se pierd este în acest caz k. În prima fază se pierd k - 1 termeni, iar în a doua fază un termen. Pierderea de

Material cu drept de autor

informație produsă de termenii lipsă afectează concluziile analizei, în special în cazul unui număr redus de observații.

☞ **Ajustarea pe baza metodei grafice**. Metoda constă în reprezentarea grafică a seriei (cronograma), pe axa **ox** (abscisa) fiind trecute momentele sau intervalele succesive de timp, iar pe axa **oy** (ordonata) înscriindu-se valorile numerice ale termenilor seriei. Se construieşte pe acelaşi grafic o dreaptă sau curbă care să unească cele două puncte extreme ale seriei cronologice astfel încât să prezinte abateri minime față de poziția valorilor reale de pe grafic.

Forma curbei astfel trasate indică legitatea matematică, forma de evoluție a fenomenului, după o dreaptă sau o funcție curbilinie.

Metoda este simplă şi rapidă, dar există pericolul interpretării subiective a graficului.

Metoda grafică precede obligatoriu aplicarea metodelor analitice de ajustare.

☞ **Ajustarea prin metoda modificării mediei absolute**. Această metodă este indicată atunci când seria cronologică prezintă tendința de creştere sub forma unei progresii aritmetice, situație evidențiată prin valorile relativ apropiate ale modificărilor absolute cu bază în lanț.

Termenii ajustați se determină cu relația:

$$Y_i = y_1 + t_i.\overline{\Delta}, \qquad i = \overline{1,n} \tag{8.28}$$

unde: $\mathbf{Y_i}$ - valorile ajustate, care înlocuiesc valorile reale; $\mathbf{y_1}$ - primul termen al seriei cronologice reale (sau un alt termen luat ca bază de ajustare); $\overline{\Delta}$ - sporul mediu (modificarea absolută medie) $\overline{\Delta} = (y_n - y_1)/(n-1)$; $\mathbf{t_i}$ - variația timpului: $\mathbf{t_1 = 0, t_2 = 1, t_3 = 2, ..., t_n = n-1}$ dacă se ia ca bază primul termen al seriei.

Primul termen ajustat este egal cu primul termen al seriei reale iar ultimul termen ajustat este egal cu ultima valoare a seriei reale:

$$Y_1 = y_1 + 0.\overline{\Delta} = y_1$$

$$Y_n = y_1 + (n-1).\overline{\Delta} = y_1 + (n-1).\frac{y_n - y_1}{n-1} = y_n$$

☞ **Ajustarea pe baza indicelui mediu de dinamică**. Această metodă este recomandabilă în situațiile în care seria cronologică are forma unei progresii geometrice cu rația $\bar{I}$. În acest caz indicii de dinamică cu bază în lanț au valori relativ apropiate.

Termenii seriei ajustate se determină cu relația:

$$Y_i = y_1.\bar{I}^{t_i} \tag{8.29}$$

Material cu drept de autor

unde: Y_i - valorile ajustate; y_1 - primul termen al seriei cronologice reale; $\bar{I}$ - indicele mediu de dinamică; $\bar{I} = \sqrt[n-1]{\frac{y_n}{y_1}}$; t_i - factorul timp; $t_i = 0, 1, 2, ..., n-1$.

Primul şi ultimul termen ajustat sunt egali cu termenii corespunzători din seria reală.

Metoda modificării absolute medii şi metoda indicelui mediu de dinamică sunt simple şi rapide, dar nu iau în calcul toate valorile absolute ale termenilor seriei.

☞ **Ajustarea prin metode analitice**. Metodele analitice de estimare a tendinței **se bazează pe folosirea funcțiilor matematice**. Alegerea funcției de ajustare se face pe baza analizei graficului şi a indicatorilor seriei cronologice. Situațiile cele mai frecvent întâlnite sunt:

* fenomenul evoluează după o **funcție liniară** atunci când graficul arată o tendință de creştere absolută constantă şi modificările cu bază în lanț au valori apropiate;

* fenomenul evoluează după o **funcție exponențială** atunci când graficul arată o tendință de creştere relativă constantă şi se obțin valori apropiate ale indicilor cu bază în lanț;

* fenomenul evoluează după o **parabolă** atunci când graficul are punct de maxim sau de minim iar diferențele dintre modificările succesive cu bază în lanț (numite modificări cu bază în lanț de ordinul doi) au valori apropiate; frecvent, pe grafic, se evidențiază numai fragmente de parabolă.

După ce se alege forma cea mai potrivită pentru funcția de ajustare, se determină parametrii prin intermediul metodei celor mai mici pătrate. Această metodă porneşte de la condiția minimizării pătratelor abaterilor valorilor ajustate (Y_i) de la valorile reale (y_i):

$$\Sigma(y_i - Y_i)^2 = \min$$

În cazul funcției liniare $Y_i = a + b.t_i$ şi condiția anterioară devine:

$$\Sigma[(y_i - (a + b.t_i)]^2 = \min$$

Din condițiile de anulare a celor două derivate parțiale,în raport cu **a** şi **b**,ale expresiei precedente rezultă sistemul de ecuații:

$$\begin{cases} n.a + b.\Sigma t_i = \Sigma y_i \\ a.\Sigma t_i + b.\Sigma t_i^2 = \Sigma t_i.y_i \end{cases}$$

Cunoscând valorile parametrilor **a** şi **b**, se pot calcula în continuare valorile ajustate Y_i.

Verificarea calculelor se face pe baza relației: $\Sigma Y_i = \Sigma y_i$.

În rezolvarea sistemului, pentru simplificare, s-a recurs la ipoteza $\Sigma t_i = 0$. Pentru aceasta, este necesar să se măsoare într-un anumit mod variația timpului, tabelul 8.5:

Material cu drept de autor

• în cazul seriilor cu **număr impar de termeni**, **t = 0** pentru termenul median, celelalte valori ale timpului fiind plasate simetric (negativ şi pozitiv) față de origine;

• în cazul seriilor cu **număr par de termeni**, celor doi termeni centrali le corespund valorile **-1** şi **+1** pe axa timpului, restul valorilor ti (pozitive şi negative) fiind de asemenea plasate simetric.

Tabelul 8.5

Variația timpului

Serii impare		Serii pare	
Anii (lunile)	t_i	Anii (lunile)	t_i
1	-3	1	-5
2	-2	2	-3
3	-1	3	-1
4	0	4	+1
5	+1	5	+3
6	+2	6	+5
7	+3	-	-
Total	0	Total	0

☞ **Analiza calității estimării tendinței**. Ajustarea seriilor cronologice se poate face prin metode diferite, iar în cazul ajustării analitice se pot utiliza diverse funcții matematice.

Alegerea celei mai bune metode de ajustare din cele disponibile presupune compararea rezultatelor obținute prin procedee diferite.

1. O primă posibilitate de comparație se bazează pe **reprezentarea grafică** a valorilor ajustate şi a celor reale. Prin compararea alurei graficelor valorilor ajustate (obținute prin diverse metode) cu graficul valorilor efective se decide care este varianta cea mai apropiată de realitate.

2. O altă metodă de apreciere a calității ajustării constă în compararea **sumei valorilor reale cu suma valorilor ajustate**: $\Sigma y_i = \Sigma Y_i$. Se alege varianta pentru care suma valorilor ajustate se află la distanță minimă de suma valorilor empirice.

3. Măsurarea obiectivă a calității ajustării se poate face şi mai exact pe baza **coeficientului de variație a valorilor ajustate de la valorile reale**. Acest indicator se calculează pentru fiecare metodă de ajustare folosită, ca raport între abaterea medie liniară a valorilor reale de la valorile ajustate şi media valorilor reale, conform relației:

$$v = \frac{\Sigma |y_i - Y_i|}{n.\bar{y}}$$

Coeficientul de variație cu valoarea cea mai mică indică cea mai bună metodă sau funcție de ajustare.

Material cu drept de autor

4. Se mai poate calcula suma pătratelor abaterilor valorilor ajustate de la cele reale $\Sigma(y_i - Y_i)^2$, alegându-se metoda de ajustare pentru care această sumă înregistrează cea mai mică valoare.

☞ **Extrapolarea**. Un obiectiv important al analizei seriilor cronologice îl reprezintă estimarea evoluției probabile în viitor a fenomenului analizat.

Extrapolarea reprezintă o prelungire a seriei cronologice în viitor, pe baza trendului observat din analiza perioadei anterioare.

Mărimile obținute prin extrapolare sunt valori probabile, orientative. Nu se poate face o predicție exactă a viitorului din mai multe motive:

• pe lângă trendul pe baza căruia se face previziunea acționează şi factori aleatori, întâmplători, care influențează nivelul real al fenomenului analizat;

• factorii de influență evidențiați prin analiza seriei cronologice îşi pot modifica acțiunea în viitor;

• există anumite limite (minime sau maxime) în evoluția fenomenelor. Aceste restricții impun o analiză critică, calitativă a rezultatelor obținute prin extrapolare.

Metodele de extrapolare sunt similare celor utilizate pentru estimarea trendului. Diferența constă în perioada de timp implicată în calcule.

Astfel, dacă analiza seriei cronologice a relevat o tendință de creştere constantă, extrapolarea se poate face prin **metoda modificării absolute medii**, conform relației:

$$Y_j = y_1 + t_j.\overline{\Delta}$$

unde: Y_j - valorile viitoare ale seriei cronologice; y_1 - un termen al seriei cronologice reale (de regulă primul); Δ - modificarea medie absolută; t_j - valorile viitoare ale factorului timp; $t_j = n, n+1, n+2, ..., n-1+m$; m - numărul intervalelor de timp pentru care se face extrapolarea.

Atunci când evoluția seriei cronologice tinde către o exponențială, extrapolarea se poate face pe baza **indicelui mediu de dinamică**:

$$Y_j = y_1.\bar{I}^{t_j}$$

unde: $\bar{I}$ - indicele mediu de dinamică.

Atunci când evoluția fenomenului a fost analizată cu ajutorul **metodelor analitice** şi s-a constatat că reflectă o anumită funcție matematică, extrapolarea se poate face utilizând forma funcției de ajustare. De exemplu, în cazul funcției liniare, extrapolarea se face astfel:

$$Y_j = a + b.t_j$$

Exemplu de calcul 8.2. Cifra de afaceri a unei Societăți Comerciale (SC) a evoluat în 11 ani în conformitate cu datele din tabelul următor, tabelul 8.6. Evaluarea s-a făcut, de fiecare dată, la sfârşitul anului calendaristic.

Material cu drept de autor

Tabelul 8.6

Evoluția cifrei de afaceri CA (în u.m. – unități monetare)

Anul	1	2	3	4	5	6	7	8	9	10	11
C.A.	54,5	61,4	67,9	76,2	85,5	92,8	101,0	109,9	118,8	125,5	131,5

Se cere analiza statistică a evoluției cifrei de afaceri.

Rezolvare.

După enunțul problemei, rezultă că aceasta este o serie cronologică de momente cu distanțe egale între momente şi se poate prelucra statistic asemănător seriilor cronologice de intervale (cu excepția nivelului mediu al termenilor seriei, care se determină după o formulă specială).

Analiza statistică se prezintă în tabelul 8.7.

Tabelul 8.7

y_t (CA)	t	$\Delta_{t/1}$ u.m.	$\Delta_{t/t-1}$ u.m.	$I_{t/1}$ %	$I_{t/t-1}$ %	$R_{t/1}$ %	$R_{t/t-1}$ %	$A_{t/t-1}$ um%
0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
54,5	1	-	-	-	-	-	-	-
61,4	2	6,9	6,9	112,66	112,66	12,66	12,66	0,545
67,9	3	13,4	6,5	124,59	110,59	24,59	10,59	0,614
76,42	4	21,7	8,3	139,82	112,22	39,82	12,22	0,679
85,5	5	31,0	9,3	156,88	112,20	56,88	12,20	0,762
92,8	6	38,3	7,3	170,27	108,54	70,27	8,54	0,855
101,0	7	46,5	8,2	185,32	108,84	85,32	8,84	0,928
109,9	8	55,4	8,9	201,65	108,81	101,65	8,81	1,010
118,8	9	64,3	8,9	217,98	108,10	117,98	8,10	1,099
125,5	10	71	6,7	230,27	105,64	130,27	5,64	1,188
131,5	11	77	6	241,28	104,78	141,28	4,78	1,255

În tabelul 8.7 se prezintă: în coloana 0 – valorile individuale absolute, de unde rezultă indicele agregat Σy_t **= 932** u.m.; în coloana 1 se prezintă indicele de momente t; în coloanele 2 şi 3 se prezintă modificarea absolută a fenomenului economic raportat la bază fixă şi bază mobilă; în coloanele 4 şi 5 se prezintă indicatori relativi raportați la bază fixă şi bază mobilă; în coloanele 6 şi 7 se prezintă ritmul de dinamică raportat la bază fixă şi bază mobilă; în coloana 8 se prezintă valoarea absolută a unui procent de creştere a fenomenului economic analizat. Calculul acestora nu ridică probleme deosebite – utilizarea expresiilor de calcul menționate anterior este simplă. De aceea, vom trece la calculul indicatorilor medii.

Indicatorii medii ai acestei serii cronologice sunt:

Material cu drept de autor

• **nivelul mediu al seriei**, calculat cu formula mediei cronologice simple deoarece aceasta este o serie cronologică de momente:

$$\bar{y} = \frac{y_t}{n-1} = \frac{\frac{54,5}{2} + 61,4 + 67,9 + ... + 118,8 + 125,5 + \frac{131,5}{2}}{11-1} = \frac{932}{10} = 93,2\,u.m.$$

• **modificarea absolută medie**:

$$\overline{\Delta} = \frac{y_n - y_1}{n-1} = \frac{131,6 - 54,5}{10} = 7,7\,u.m.$$

• **indicele mediu de dinamică**:

$$\bar{I} = \sqrt[n-1]{\frac{y_n}{y_1}} = \sqrt[10]{\frac{131,5}{54,5}} = 1,0921 \Rightarrow 109,21\%$$

• **ritmul mediu de creştere**:

$$\overline{R}\% = \bar{I}\% - 100 = 109,21 - 100 = 9,21\%$$

Arată o creştere medie anuală cu 9,21% a cifrei de afaceri.

Ajustarea prin metoda modificării absolute medii ($\overline{\Delta}$) se realizează cu ajutorul relației:

$$Y_i = y_1 + t_i.\overline{\Delta}, \quad \Rightarrow \quad y_1 = 54,5\,u.m;\ \overline{\Delta} = 7,7\,u.m.$$

Valorile ajustate calculate pe baza acestei relații şi abaterile valorilor ajustate de la valorile reale sunt prezentate în tabelul 8.8.

Tabelul 8.8

Ajustarea prin metoda modificării absolute medii

t_i	Valorile reale y_i	Valorile ajustate Y_i	Abaterile $\lvert y_i - Y_i \rvert$
0	54,5	54,5	0
1	61,4	62,5	1,1
2	67,9	69,9	2,0
3	76,2	77,6	1,4
4	85,5	85,3	0,2
5	92,8	93,0	0,2
6	101,0	100,7	0,3
7	109,9	108,4	1,5
8	118,8	116,1	2,7
9	125,5	123,8	1,7
10	131,5	131,5	0
Total	-	-	9,6

Material cu drept de autor

Ajustarea prin metoda indicelui mediu de dinamică ($\bar{I}$) se bazează pe relația:

$$Y_i = y_1 \times \bar{I}^{t_i}, \qquad \Rightarrow y_1 = 54{,}5; \quad \bar{I} = 1{,}0921$$

Valorile ajustate conform acestei metode şi abaterile de la mărimile reale y_i sunt prezentate în tabelul 8.9.

Tabelul 8.9

Ajustarea prin metoda indicelui mediu de dinamică

t_i	Valorile reale y_i	Valorile ajustate Y_i	Abaterile $\lvert y_i - Y_i \rvert$
0	54,5	54,5	0
1	61,4	59,5	1,9
2	67,9	65,0	2,0
3	76,2	71,0	5,2
4	85,5	77,5	8,0
5	92,8	84,7	8,1
6	101,0	92,5	9,0
7	109,9	101,0	8,9
8	118,8	110,3	8,5
9	125,5	120,4	5,1
10	131,5	131,5	0
Total	1025,0	967,9	57,6

Ajustarea prin metode analitice, folosind trendul liniar (vezi tabelul 8.10) se bazează pe relația:

$$Y_i = a + b.t_i, \Rightarrow a = \frac{\Sigma y_i}{n} = \frac{1025}{11} = 93{,}18; \; b = \frac{\Sigma t_i y_i}{\Sigma t_i^2} = \frac{877}{110} = 7{,}97$$

Aşadar:

$$Y_i = 93{,}18 + 7{,}97.t_i$$

Tabelul 8.10

Ajustarea analitică cu trend liniar

t_i	y_i	$y_i.t_i$	t_i^2	$Y_i = 93{,}18 + 7{,}97.t_i$	$\lvert y_i - Y_i \rvert$
-5	54,5	-272,5	25	53,33	0,95
-4	61,4	-245,6	16	61,30	0,10
-3	67,9	-203,7	9	69,27	1,37
-2	76,2	-152,4	4	77,24	1,04

Material cu drept de autor

-1	85,5	-85,5	1	85,21	0,29
0	92,8	0	0	93,18	0,38
1	101,0	101,0	1	101,15	0,15
2	109,9	219,8	4	109,12	0,78
3	118,8	356,4	9	117,09	1,71
4	125,5	502,0	16	125,06	0,44
5	131,5	657,5	25	133,03	1,53
Total	1025	877	110	1024,98	8,74

Ajustarea prin metode analitice, folosind trend parabolic:

$$Y_i = a + b.t_i + c.t_i^2$$

unde: a,b,c - parametrii funcției parabolice de gradul doi.

Prin aplicarea metodei celor mai mici pătrate rezultă sistemul de ecuații normale:

$$\begin{cases} n.a + b.\Sigma t_i + c.\Sigma t_i^2 = \Sigma y_i \\ a.\Sigma t_i + b.\Sigma t_i^2 + c.\Sigma t_i^3 = \Sigma t_i y_i \\ a.\Sigma t_i^2 + b.\Sigma t_i^3 + c.\Sigma t_i^4 = \Sigma t_i^2 y_i \end{cases}$$

Întrucât $\Sigma t_i = 0$, rezultă:

$$\begin{cases} n.a + c.\Sigma t_i^2 = \Sigma y_i \\ b.\Sigma t_i^2 = \Sigma t_i y_i \\ a.\Sigma t_i^2 + c.\Sigma t_i^4 = \Sigma t_i^2 y_i \end{cases}$$

Din a doua ecuație a sistemului se obține:

$$b = \frac{\Sigma t_i y_i}{\Sigma t_i^2} = \frac{877}{110} = 7,97$$

Tabelul 8.11

Ajustarea analitică cu trend parabolic

t_i	y_i	t_i^4	$t_i^2.y_i$	$Y_i = 54,13 - 0,5.t_i - 0,057.t_i^2$	$\lvert y_i - Y_i \rvert$
-5	54,5	625	1362,5	53,36	1,14
-4	61,4	256	982,4	61,31	0,09
-3	67,9	81	611,1	69,27	1,37
-2	76,2	16	304,8	77,23	1,03
-1	85,5	1	85,5	85,19	0,31

Material cu drept de autor

0	92,8	0	0	93,16	0,30
1	101,0	1	101	101,13	0,13
2	109,9	16	439,6	109,11	0,79
3	118,8	81	1069,2	117,09	1,71
4	125,5	256	2008	125,07	0,43
5	131,5	625	3287,5	133,06	1,56
Total	1025	1958	10251,6	1024,98	8,92

Conform calculelor prezentate în tabelul 8.11, sistemul anterior devine:

$$\begin{cases} 11.a + 110.c = 1025 \\ 110.a + 1958.c = 10251{,}6 \end{cases} \Rightarrow a = 93{,}16;\ c = 0{,}0019$$

Funcția de ajustare este:

$$Y_i = 93{,}16 + 7{,}97.t_i + 0{,}0019.t_i^2$$

Alegerea celei mai potrivite funcții de ajustare se poate face pe baza **coeficientului de variație** a valorilor ajustate față de cele reale:

$$\nu = \frac{\Sigma|y_i - Y_i|}{n.\bar{y}},$$

calculat pentru toate metodele de ajustare utilizate.

Ajustarea după metoda modificării absolute medii:

$$\nu_1 = \frac{9{,}6}{11.93{,}2} = 0{,}00936$$

Ajustarea după metoda indicelui mediu de dinamică:

$$\nu_2 = \frac{57{,}6}{11.93{,}2} = 0{,}05618$$

Ajustarea prin metode analitice, cu funcție liniară:

$$\nu_3 = \frac{8{,}74}{11.93{,}2} = 0{,}00853$$

Ajustarea prin metode analitice, cu funcție parabolică de gradul doi:

$$\nu_4 = \frac{8{,}92}{11.93{,}2} = 0{,}00870$$

Cel mai mic coeficient de variație se înregistrează în cazul ajustării după o funcție liniară.

Prognoza evoluției cifrei de afaceri a societății comerciale, pentru următorii doi ani, se va face după funcția liniară, întrucât este cea mai bună

Material cu drept de autor

metodă de estimare a trendului în această aplicație. Seria cronologică va fi prelungită în viitor dând variabilei ti valorile următoare pe axa timpului: 6 şi 7.

$$Y_6 = 93{,}18 + 7{,}97 \times 6 = 141 u.m.$$

$$Y_7 = 93{,}18 + 7{,}97 \times 7 = 148{,}97 u.m.$$

8.6. Analiza statistică a variațiilor sezoniere

În manifestarea concretă a fenomenelor economice există unele variații cu caracter de regularitate, legate de succesiunea anotimpurilor.

Variațiile sezoniere se manifestă în producția şi desfacerea unor bunuri de consum, în turism, construcții, transport fluvial, agricultură etc.

Factorul sezonier poate fi izolat prin eliminarea din cadrul seriei cronologice a trendului şi a abaterilor aleatoare. Este importantă cunoaşterea periodicității producerii fenomenului.

Cunoaşterea gradului de sezonalitate este importantă în deciziile de producție şi desfacere din anumite domenii de activitate.

Sezonalitatea este una dintre componentele seriilor cronologice. În cadrul seriilor cronologice suficient de lungi pot fi evidențiate mai multe componente:

• **trendul** (tendința centrală) este componenta esențială a seriilor cronologice şi exprimă tendința generală de evoluție pe termen lung; este rezultatul acțiunii sistematice, constante a factorilor de influență esențiali;

• **sezonalitatea** se manifestă sub forma unor oscilații periodice de la tendința generală, care se succed la intervale constante, mai mici de un an; aceste variații au caracter repetitiv şi sunt produse de factori naturali care condiționează activitatea din construcții, agricultură, transport fluvial, turism etc.

• **ciclicitatea** reprezintă o altă formă de osculații periodice de la trend, dar acestea se repetă la intervale neegale de timp, de lungă durată (mai mari de un an); ciclurile macroeconomice produse de modificările în eficiența utilizării factorilor de producție sunt cele mai cunoscute exemple din această categorie;

• **componenta aleatoare** este rezultatul acțiunii factorilor întâmplători, accidentali şi se manifestă prin abateri imprevizibile de la trend.

În cadrul seriilor cronologice, componentele prezentate se combină folosind modelul aditiv sau modelul multiplicativ. Alegerea modelului se face în funcție de modul cum reacționează oscilațiile sezoniere față de componenta de tendință.

Dacă oscilațiile sezoniere rămân constante indiferent dacă fenomenul manifestă o tendință de creştere sau de scădere, se foloseşte modelul aditiv. În cazul modelului aditiv, toate componentele sunt exprimate în aceleaşi unități de măsură, componentele de sezonalitate, ciclicitate şi cele aleatoare reprezentând abateri față de componenta de tendință.

Dacă oscilațiile sezoniere se modifică odată cu componenta de tendință, se aplică modelul multiplicativ. În acest caz, numai componenta de

Material cu drept de autor

tendință se exprimă în unitățile de măsură ale fenomenului analizat, celelalte componente exprimându-se sub formă de coeficient sau procentual.

Măsurarea oscilațiilor sezoniere presupune, în prealabil, ajustarea seriei prin metoda mediilor mobile sau o metodă analitică adecvată obținându-se valorile ajustate y_{ij}.

În cazul modelului multiplicativ, pentru măsurarea sezonalității se calculează indicii de sezonalitate astfel:

1. Se calculează rapoartele dintre valorile reale y_{ij} şi valorile ajustate $\hat{y}_{ij}$. Deoarece valorile inițiale y_{ij} conțin trendul, componenta aleatoare şi sezonalitatea, din calcul rezultă produsul dintre componenta sezonieră şi cea aleatoare(reziduală):

$$\frac{y_{ij}}{\hat{y}_{ij}} = S_j . R_{ij}$$

unde: y_{ij} - valorile reale ale seriei cronologice; $\hat{y}_{ij}$ - valorile ajustate, trendul; S_j - componenta sezonieră; R_{ij} - componenta aleatoare; i - anul; j - sezonul (trimestru).

2. Pentru a elimina efectul factorilor întâmplători, din rapoartele determinate anterior se calculează medii aritmetice parțiale pe sezoane, care reprezintă estimatorii bruți ai componentei sezoniere:

$$S'_j = \frac{\sum_{i=1}^{n} \frac{y_{ij}}{\hat{y}_{ij}}}{n}$$

Dacă produsul estimatorilor componentei sezoniere este egal cu unu, înseamnă că aceştia sunt chiar indicii de sezonalitate şi se trece la pasul 4. În caz contrar, se trece la pasul 3 pentru corectarea acestora.

3. Se calculează media estimatorilor bruți ai componentei sezoniere şi apoi se calculează raportul dintre fiecare estimator şi media acestora, obținându-se indicii de sezonalitate:

$$\bar{S} = \frac{\sum_{j=1}^{m} S'_j}{m} \quad şi \quad I_{S_j} = \frac{S'_j}{\bar{S}}$$

4. Se calculează valorile desezonalizate raportând valorile reale la indicii de sezonalitate corespunzători.

Aplicarea **modelului aditiv** presupune parcurgerea următoarelor etape:

1. Se calculează diferențele dintre valorile reale y_{ij} şi valorile ajustate $\hat{y}_{ij}$ şi rezultă suma componentei sezoniere cu cea reziduală:

$$y_{ij} - \hat{y}_{ij} = S_j + R_{ij} ,$$

2. Se calculează media diferențelor pe fiecare sezon **j**, obținându-se o estimare a componentei sezoniere:

Material cu drept de autor

$$S'_j = \frac{\sum_{i=1}^{n}(y_{ij} - \hat{y}_{ij})}{n}$$

Dacă suma estimatorilor componentei sezoniere este diferită de zero, atunci se trece la pasul 3.

3. Dacă abaterile sezoniere sunt comparabile în valoare absolută, se diminuează componentele sezoniere S'_j cu media lor, obținându-se abaterile sezoniere corectate:

$$S_j = S'_j - \frac{\sum_{j=1}^{m} S'_j}{m}$$

unde: $\mathbf{S_j}$ - abaterile sezoniere corectate; **j** - sezonul, **j = 1, ..., m**.

În caz contrar, se repartizează, proporțional, pe sezoane, diferența obținută la pasul precedent.

4. Se corectează seria cronologică inițială, eliminându-se influența factorului sezonier prin scăderea abaterilor sezoniere corectate $\mathbf{S_j}$ din toți termenii $\mathbf{y_{ij}}$:

$$y_{ij} - S_j, \quad i = 1,\ldots,n \;\; şi \;\; j = 1,\ldots,m$$

Termenii seriei astfel corectate conțin trendul şi abaterea aleatoare.

Indiferent de modelul de combinare a componentelor folosit, datele desezonalizate se ajustează printr-o metodă analitică corespunzătoare şi pe baza ecuației de ajustare se poate extrapola componenta de tendință prelungind variabila timp pentru orizontul de prognoză şi păstrând parametrii ecuației de ajustare nemodificați dacă nu se întrevăd modificări semnificative în evoluția fenomenului.

Pentru previzionarea seriilor care prezintă oscilații sezoniere, valorile previzionate după tendință se sezonalizează prin înmulțire cu coeficienții de sezonalitate corespunzători în cazul modelului multiplicativ sau prin adunarea abaterilor sezoniere corespunzătoare în cazul modelului aditiv.

Pentru exemplificare se folosesc datele următoare referitoare la producția trimestrială de înghețată a unei societăți comerciale.

Exemplu de calcul 8.3. Producția de înghețată raportată de o societate comercială în trei ani consecutivi A, B C, pe trimestre, în unități monetare (u.m.) echivalente, este prezentată în tabelul următor (date convenționale), tabelul 8.12:

Tabelul 8.12. Date inițiale

Anul	A				B				C			
Trim.	I	II	III	IV	I	II	III	IV	I	II	III	IV
u.m.	20	70	150	40	30	90	180	60	40	110	240	110

Material cu drept de autor

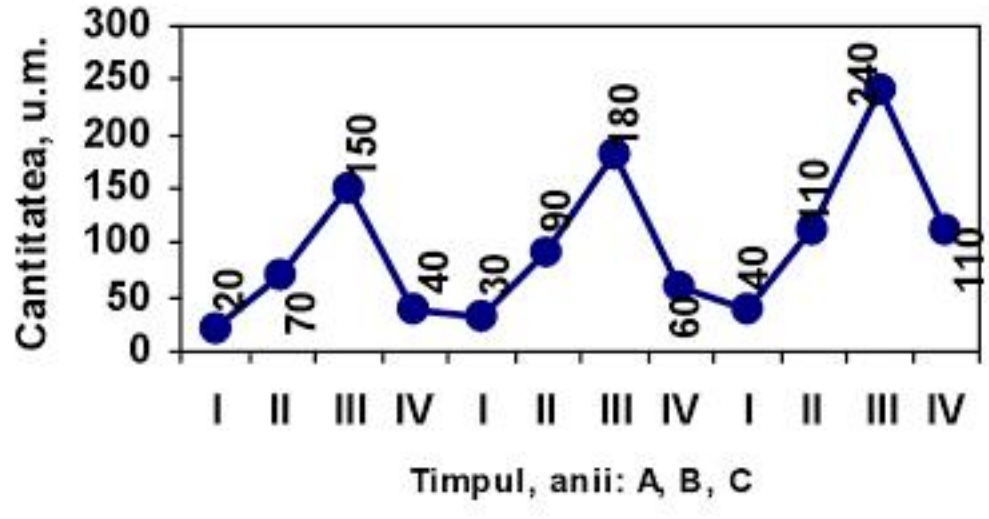
Se cere să se determine tendința pe care a înregistrat-o producția de înghețată pe perioada specificată.

Rezolvare.

a. Se face reprezentarea grafică din care rezultă oscilațiile sezoniere, cât şi tendința de creştere a producției de bere.

b. Se face desezonalizarea seriei prin calculul mediilor mobile din câte patru termeni, operând de la început deplasarea cu jumătate de termen spre dreapta.

Se construieşte cronograma pe baza datelor inițiale ale seriei statistice.



Pentru ajustarea seriei cronologice se foloseşte metoda mediilor mobile. Cum periodicitatea termenilor seriei este trimestrială, se calculează medii mobile din patru termeni care, pentru a fi centrate, se calculează în două faze: medii mobile provizorii şi medii mobile definitive.

Tabelul 8.13. Calcule intermediare

Anul	Trimestrul j	Producția y_{ij}	Trend $\hat{y}_{ij}$	$\frac{y_{ij}}{\hat{y}_{ij}}$
A	**B**	**1**	**2**	**3**
A	I	20	-	-
	II	70	-	-
	III	150	71,25	2,10
	IV	40	75	0,53
B	I	30	81,25	0,37
	II	90	87,5	1,03
	III	180	91,25	1,97
	IV	60	95	0,63
C	I	40	105	0,38
	II	110	118,75	0,93
	III	240	-	-
	IV	110	-	-

Material cu drept de autor

Mediile mobile provizorii:

$$\bar{Y}_1 = \frac{20+70+150+40}{4} = 70 \qquad \bar{Y}_2 = \frac{70+150+40+30}{4} = 72{,}5$$

$$\bar{Y}_3 = \frac{150+40+30+90}{4} = 77{,}5 \qquad \bar{Y}_4 = \frac{40+30+90+180}{4} = 85$$

$$\bar{Y}_5 = \frac{30+90+180+60}{4} = 90 \qquad \bar{Y}_6 = \frac{90+180+60+40}{4} = 92{,}5$$

$$\bar{Y}_7 = \frac{180+60+40+110}{4} = 125 \quad \bar{Y}_8 = \frac{60+40+110+240}{4} = 112{,}5$$

$$\bar{Y}_9 = \frac{40+110+240+110}{4} = 125$$

Mediile mobile definitive calculate ca medii mobile de câte două medii mobile provizorii sunt prezentate în tabelul 8.13, coloana 2.

Graficul reprezentat de corelograma de mai sus sugerează utilizarea modelului multiplicativ. Pentru înlăturarea componentei de tendință se calculează raportul între fiecare termen real şi cel ajustat, Tabelul 8.13, coloana 3.

Indicii de sezonalitate se calculează ca raport între fiecare medie trimestrială şi media mediilor trimestriale, tabelul 8.14, coloana 5.

Tabelul 8.14. Calculul indicilor de sezonalitate.

Trimestru	$y_{ij}/\hat{y}_{ij}$			Indici de sezonalitate, bruți S_j	Indici de sezonalitate I_{s_j}
	A	B	C		
0	1	2	3	4	5
I	-	0,37	0,38	0,375	0,378
II	-	1,03	0,93	0,98	0,989
III	2,10	1,97	-	2,03	2,048
IV	0,53	0,63	-	0,58	0,585

Dacă se doreşte previzionarea valorilor pentru anul următor este necesar să se calculeze valorile desezonalizate, tabelul 8.15.

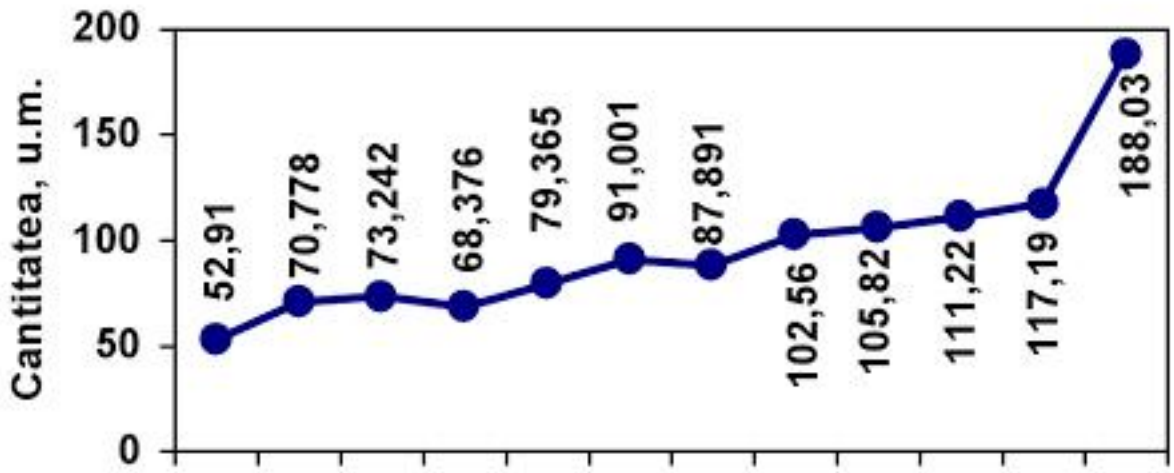
Tabelul 8.15. Analiza sezonalității producției trimestriale.

Anul	Trimestrul j	Producția y_{ij}	Indici de sezonalitate I_{s_j}	Valori desezonalizate, col.1/col.2
A	**B**	**1**	**2**	**3**
A	I	20	0,378	52,910
	II	70	0,989	70,778
	III	150	2,048	73,242
	IV	40	0,585	68,376

Material cu drept de autor

B	I	30	0,378	79,365
	II	90	0,989	91,001
	III	180	2,048	87,891
	IV	60	0,585	102,564
C	I	40	0,378	105,820
	II	110	0,989	111,223
	III	240	2,048	117,187
	IV	110	0,585	188,034
Total	-	-	-	1048,391

Se reprezintă grafic corelograma valorilor desezonalizate.



Graficul sugerează utilizarea modelului liniar pentru ajustare:

$$Y_i = a + b.t_i$$

Tabelul 8.16.

Anul	Trim. j	Valori dese-zonalizate, (y_i)	t_i	t_i^2	$t_i y_i$
A	**B**	**0**	**1**	**2**	**3**
A	I	52,910	-11	121	-582,01
	II	70,778	-9	91	-637,002
	III	73,242	-7	49	-512,694
	IV	68,376	-5	25	-341,880
B	I	79,365	-3	9	-238,095
	II	91,001	-1	1	-91,001
	III	87,891	1	1	87,891
	IV	102,564	3	9	307,692

Material cu drept de autor

C	I	105,820	5	25	529,100
	II	111,223	7	49	778,561
	III	117,187	9	81	1054,687
	IV	188,034	11	121	2068,374
Total	-	1048,391	0	582	2423,623

Pentru rezolvarea modelului liniar se foloseşte sistemul ecuațiilor normale:

$$\begin{cases} n.a + b\Sigma t_i = \Sigma y_i \\ a\Sigma t_i + b\Sigma t_i^2 = \Sigma t_i y_i \end{cases}$$

Pentru $\Sigma t_i = 0,$ sistemul de ecuații normale devine:

$$\begin{cases} n.a = \Sigma y_i \\ b\Sigma t_i^2 = \Sigma t_i y_i \end{cases} \Rightarrow \begin{cases} a = \dfrac{\Sigma y_i}{n} \\ b = \dfrac{\Sigma t_i y_i}{\Sigma t_i^2} \end{cases}$$

Pentru calcule s-a întocmit tabelul 8.16. Rezultă

$$a = \frac{\Sigma y_i}{n} = \frac{1048,391}{12} = 87,36 \quad b = \frac{2423,623}{582} = 4,24$$

$$Y_i = 87,36 + 4,24.t_i$$

Pentru determinarea valorilor previzionate pentru anul următor D se prelungeşte variabila timp, se calculează valorile de tendință teoretice care se înmulțesc cu coeficienții de sezonaliate, tabelul 8.17.

Tabelul 8.17. Valori previzionate pentru anul D

Anul	Trim. j	t_i	$Y_i = 87,36 + 4,24.t_i$	Indici de sezonalitate I_{S_j}	Valori previzionate
D	I	13	142,48	0,378	53,84
	II	15	150,96	0,989	149,30
	III	17	159,44	2,048	326,53
	IV	19	167,92	0,585	98,23

Material cu drept de autor

CAP.9. METODA INDICILOR

9.1. Conceptul de indici statistici

Prin categoria statistică a indicilor se înțelege o expresie de un anumit fel a mărimilor relative, având funcția de a măsura schimbarea medie în timp – aspectul dinamic – al fenomenelor şi proceselor social-economice sau de a le compara în spațiu, la un moment dat sau într-o anumită perioadă – aspectul static.

Procedeul de construire a oricărui indice rezultă dintr-o relație matematică elementară: se construieşte un raport prin care anumite date statistice luate în studiu se compară cu alte date, cu caracter analog, dintr-o altă perioadă de timp sau din aceeaşi perioadă, însă, din spațiu diferit. Datele luate în studiu – **date raportate** – apar la numărătorul raportului, iar datele cu care se face comparația – **baza de raportare** – apar la numitorul acestuia.

Pe cât de simplă este schema matematică a oricărui indice, pe atât de complexe sunt, pe de altă parte, conținutul şi metodologia de calcul al indicelui, ca procedeu de analiză statistică a fenomenelor şi proceselor social-economice, deci, ca metodă specifică a statisticii în înțelesul ei clasic de ştiință socială, fundamentată pe analiza ştiințifică a realității. De aceea, problema centrală care se pune atunci când se elaborează un indice economic constă în a înțelege conținutul, esența fenomenului economic pentru care se calculează.

Faptul că prin metoda indicelui suntem în măsură să caracterizăm statistic esența fenomenelor studiate are un înțeles teoretic precis. În general, datele luate în studiu (cele raportate, care figurează la numărătorul raportului), caracterizează un static al fenomenului respectiv. Dar, de îndată ce le apreciem comparativ, în raport cu datele dintr-o perioadă aleasă ca bază şi care reprezintă o altă etapă a dezvoltării fenomenului studiat, noi relevăm aspectul dinamic al dezvoltării, mişcarea şi schimbarea fenomenului în timp. Tocmai în aceasta constă sensul principal şi semnificația comparării prin metoda indicilor.

Compararea datelor curente cu datele perioadei de bază ne oferă posibilitatea de a sesiza noul pe care-l cuprind, de a măsura şi caracteriza proporțiile schimbărilor care au survenit şi nivelul de dezvoltare atins, precum şi de a releva calitatea specifică a fenomenului studiat.

Din punctul de vedere al sferei de cuprindere se disting două categorii de indici: ***indici individuali***, exprimând raportul de mărime – în timp sau spațiu – al unui fenomen (element) oarecare dintr-o colectivitate statistică şi ***indici generali*** sau ***de grup***, exprimând schimbări medii ale întregii colectivități de fenomene (obiecte) sau ale caracteristicilor unor grupe de fenomene – părți ale colectivității – supuse studiului.

Din definiția dată indicilor statistici, rezultă că aceştia reprezintă o mărime relativă, adimensională, având anumite particularități care îl deosebesc de alte mărimi relative. Aceste particularități apar îndeosebi atunci când se

Material cu drept de autor

calculează indici generali, de grup, sau indici analitici, caracterizând dinamica medie a unor fenomene complexe cu două sau mai multe caracteristici, neînsumabile direct. Tocmai de aceea noțiunea de indice se referă – în sensul ei strict tehnic – la indice de grup.

La construirea indicilor de grup se pun o serie de probleme legate de măsurarea obiectelor luate în studiu.

Astfel, pentru a compara dinamic sau sincronic producția industrială este necesar, în prealabil, să se rezolve corect metoda de determinare a volumului producției. Or, volumul total al producției industriale este un fenomen complex, în compunerea căruia intră o serie de elemente (de exemplu, o tonă de cărbune, o tonă de zahăr etc.) ce nu se pot măsura cu aceeaşi unitate de măsură.

La elaborarea indicilor se pune, de aceea, problema întrunirii lor într-un singur volum, adoptându-se în acest scop, drept unitate de măsură, o caracteristică ce le este comună. În teoria şi practica statisticii, valoarea diferitelor produse – îndeosebi valoarea lor în expresie bănească – a fost adoptată ca unitate de măsură comună.

9.2. Indici în formă de bază şi în formă de lanț

Una din problemele principale ce se pun la construirea indicilor este problema **bazei de comparație**.

Construirea indicelui, luându-se drept bază de comparație o perioadă anterioară, prezintă interes din punctul de vedere al evaluării ritmului de dezvoltare al fenomenului analizat.

Dar, pentru construirea indicilor se mai poate lua şi o altă bază de comparație şi anume – nivelul aceluiaşi fenomen din alt loc al spațiului (din altă unitate economică, altă zonă geografică etc.), care pune probleme specifice ale comparării pe plan interzonal a fenomenului analizat.

În funcție de perioada aleasă drept bază de comparație se disting două feluri de indici: ***indici în formă de bază***, când compararea datelor pe diferite perioade se face în raport cu una şi aceeaşi perioadă de bază şi ***indici în formă de lanț*** (cu bază variabilă), când compararea datelor privind nivelul unor indicatori economici pe diferite perioade se face în raport cu datele corespunzătoare ale perioadei precedente.

Dacă se notează, de exemplu, cantitățile unui anumit produs, fabricate în zece ani succesivi prin:

$$q_0, q_1, q_2, \ldots, q_9$$

atunci seria indicilor în formă de bază este:

$$\frac{q_1}{q_0}; \frac{q_2}{q_0}; \ldots; \frac{q_9}{q_0} \qquad (9.1)$$

iar a indicilor în formă de lanț:

Material cu drept de autor

$$\frac{q_1}{q_0};\frac{q_2}{q_1};\ldots;\frac{q_9}{q_8} \qquad (9.2)$$

În expresiile (9.1) şi (9.2) este vorba de indici individuali exprimați în aceleaşi unități de măsură.

O proprietate remarcabilă a indicilor individuali în formă de lanț constă în aceea că din serii de asemenea indici se pot construi serii corespunzătoare de indici de bază. Acest lucru se obține făcând produsul succesiv al indicilor în lanț.

De exemplu, dacă se consideră produsul indicilor în formă de lanț pe primii cinci ani, folosind notația de mai sus, se obține indicele de bază al celui de al cincilea an, adică raportul:

$$i_{4/0} = \frac{q_4}{q_0}$$

Pentru indicii generali, de grup, această proprietate se menține numai dacă este asigurată pe deplin comparabilitatea datelor, folosindu-se ponderi constante la elaborarea indicilor de grup în lanț.

Tabelul 9.1

Dinamica productivității muncii, în unități naturale,
la unele produse industriale, pe patru ani

Indicatori	Indici în formă de bază				Indici în formă de lanț			
	I	II	III	IV	I	II	III	IV
0	1	2	3	4	5	6	7	8
Producția de energie electrică pe muncitor	113	140	155	167	113	124,0	110,7	107,7
Producția de cărbune pe muncitor	105	111	119	130	105	105,7	107,2	109,2
Producția extracției de țiței pe muncitor	104	107	109	111	104	102,8	102,8	101,8
Producția de ciment pe muncitor	112	123	132	145	112	109,8	107,3	109,8

Legătura dintre indicii în formă de lanț şi indicii în formă de bază se poate exprima şi sub o altă formă – inversă şi anume: împărțind indicele în formă de bază al perioadei **k** la indicele în formă de bază al perioadei **k-1**, se obține indicele în formă de lanț ale perioadei **k**, față de perioada **k-1**.

Astfel, considerând indicii în formă de bază ai ultimilor patru ani din exemplu de mai sus şi făcând rapoartele respective, în sensul arătat, se obține şirul respectiv al indicilor în formă de lanț:

$$\frac{q_7}{q_0}:\frac{q_6}{q_0}=\frac{q_7}{q_6} \qquad \frac{q_8}{q_0}:\frac{q_7}{q_0}=\frac{q_8}{q_7} \qquad \frac{q_9}{q_0}:\frac{q_8}{q_0}=\frac{q_9}{q_8}$$

Material cu drept de autor

Indicii în formă de bază sunt instrumente prețioase pentru a caracteriza rezultatele activității economice pe perioade de timp mai mari, spre deosebire de indicii în formă de lanț care ilustrează schimbarea nivelului unui fenomen sau proces economic în perioada considerată, față de cea precedentă.

Se ilustrează semnificația indicilor în formă de bază şi în formă de lanț printr-un exemplu prezentat în tabelul 9.1.

Dacă se examinează datele din acest tabel ne convingem uşor că fiecare din seriile indicilor individuali – de bază şi de lanț – are un sens anumit: indicii de bază ilustrează realizările obținute în decursul întregii perioade ce s-a scurs de la anul luat drept bază de comparație, iar indicii în lanț – rezultatele fiecărui an în parte, față de cele din anul precedent. Aşa, de exemplu, indicele în formă de bază în anul III la producția de energie electrică pe un muncitor ne arată că în trei ani nivelul acestui indicator a crescut cu 55%, în comparație cu anul I luat ca bază.

9.3. Clasificarea indicilor după funcțiile lor cognitive

În principiu, un fenomen complex supus studiului statistic are un dublu caracter: el apare sub forma produsului a doi factori din care unul este **factor de volum (cantitativ)** pe care-l notăm cu $\mathbf{f_i}$, iar celălalt – un **factor de calitate (calitativ)** pe care-l notăm cu $\mathbf{x_i}$.

Această separare a factorilor după natura lor este necesară întrucât la construirea indicilor pentru ansamblul de elemente complexe trebuie avut în vedere faptul că valorile individuale ale factorilor înregistrați pot fi însumabile sau neînsumabile din punct de vedere economic. În unele cazuri, valorile individuale ale factorilor cantitativi pot fi însumate direct (produse de acelaşi fel - numărul angajaților, numărul tractoarelor de aceeaşi putere etc.), iar altele, nu sunt însumabile (cantitatea de produse diferite – maşinile agricole de diferite tipuri dintr-o unitate agricolă, cantitățile de alimente cumpărate de o familie într-o lună etc.). Valorile factorilor calitativi sunt întotdeauna neînsumabile direct. Deci, determinarea nivelului totalizator al valorilor luate în calculul indicilor, în cazurile în care valorile individuale nu pot fi însumate direct din punct de vedere economic, necesită folosirea ponderilor.

Ponderea are rolul de comăsurător al valorilor factorilor neînsumabili şi figurează întotdeauna în numărătorul şi numitorul raportului cu aceeaşi valoare. Ca atare, ***elementul care rămâne constant în numărător şi numitor are denumirea generică de pondere***. Deci, ponderea poate să fie atât factorul cantitativ, cât şi cel calitativ.

Variabila a cărei variație interesează figurează la numărătorul şi numitorul raportului, la unități diferite de timp sau spațiu, iar indicele de grup care rezultă se numeşte ***indice factorial***.

Deci, ***indicii statistici sunt rezultatul raportului dintre nivelurile atinse de un fenomen în două unități diferite de timp sau de spațiu***. Dacă se compară nivelul unui fenomen din perioada curentă cu cel dintr-o perioadă

Material cu drept de autor

anterioară rezultă ***indicele dinamicii***. Rezultatul comparării a aceluiaşi fenomen, în aceeaşi perioadă de timp, situat în două unităţi teritoriale, se concretizează în ***indici teritoriali*** sau ***de spaţiu***. Raportul nivelului realizat la nivelul planificat (programat) al aceluiaşi fenomen se obţin ***indicii planului***. Aceasta înseamnă că indicatorii care se prezintă la numărătorul şi numitorul indicilor au întotdeauna acelaşi conţinut, deosebindu-se în cazul indicilor dinamicii şi ai planului prin timpul la care se referă.

Conţinutul indicatorului comparat determină denumirea indicilor calculaţi. Rezultă că mărimile relative ale dinamicii, ale planului şi de coordonare pot fi considerate drept ***indici individuali***, notaţi cu ***i***, care exprimă variaţia relativă la nivelul unei singure unităţi de observare.

Indicii de grup simbolizaţi cu **I** se calculează la nivelul unei grupe sau pe întreg ansamblu şi exprimă în acelaşi timp şi variaţia medie relativă a fenomenului studiat. Ca atare, indicele de grup nu este o sumă a indicilor individuali respectivi, ci o medie a acestora. Indicii individuali fiind mărimi relative, media lor poate să fie aritmetică sau armonică.

Construirea şi folosirea indicilor de grup în caracterizarea unui fenomen complex presupune efectuarea unei analize calitative a indicatorului a cărui variaţie se studiază şi din care trebuie să rezulte: ***separarea factorilor, în cantitativi (extensivi) şi calitativi (intensivi); dacă valorile factorilor sunt însumabile direct sau nu; ce bază de comparaţie trebuie folosită; sistemul de ponderare care trebuie utilizat; ce relaţie de calcul poate fi aplicată având în vedere datele de care se dispune***.

In funcţie de aceste elemente, ***indicii de grup*** se pot construi sub formă de ***indici agregaţi***, ***indici calculaţi ca medie a indicilor individuali*** şi ca ***indici determinaţi ca raport între două medii***. La rândul lor, aceste forme de indici de grup se particularizează în funcţie de baza de comparaţie (fixă sau mobilă) şi de ponderile utilizate (constante sau variabile).

Complexitatea este caracteristică fenomenelor social-economice întrucât, acestea, se modifică sub influenţa a o serie de factori, care sunt grupaţi în factori cantitativi şi calitativi. Astfel, nivelul cheltuielilor de producţie este influenţat de costul pe unitatea de produs (factor calitativ - $\mathbf{x}_i$) şi de volumul producţiei (factor cantitativ - $\mathbf{f}_i$); fondul de salarii depinde de salariul mediu nominal (factor calitativ - $\mathbf{x}_i$) şi de numărul personalului salariat (factor cantitativ - $\mathbf{f}_i$) etc.

Vom considera spre exemplu indicii care exprimă raportul dintre nivelul realizat al unei realizări productive, faţă de nivelul planificat. Indicii care exprimă acest raport se pot construi ca indici ai sarcinilor de plan pentru o singură categorie de produse (indici individuali) sau pentru o colectivitate de produse eterogene, neînsumabile direct (indici de grup).

- indicii individuali:

$$i^{x}_{pl/0} = \frac{x_{pl}}{x_0}; \quad i^{f}_{pl/0} = \frac{f_{pl}}{f_0} \tag{9.3}$$

- indici de grup:

Material cu drept de autor

$$I^{x}_{pl/0} = \frac{\sum x_{pl} f_{pl}}{\sum x_0 f_{pl}}; \quad I^{f}_{pl/0} = \frac{\sum x_0 f_{pl}}{\sum x_0 f_0} \tag{9.4}$$

în care: $\mathbf{x_{pl}}$ şi $\mathbf{x_0}$ reprezintă nivelul planificat al factorului calitativ pentru perioada curentă (actuală) şi respectiv nivelul efectiv din perioada de bază; $\mathbf{f_{pl}}$ şi $\mathbf{f_0}$ – nivelul factorului cantitativ în aceleaşi perioade.

De remarcat că la construirea indicilor de grup se modifică numai factorul a cărui variaţie o urmărim, în timp ce factorul al doilea rămâne constant, îndeplinind rolul de pondere. ***La alcătuirea indicilor*** folosiţi în practica statisticii economice, ***factorii cantitativi se iau ca pondere la nivelul perioadei curente, iar cei calitativi la nivelul perioadei de bază***. În cazul indicilor de grup construiţi mai sus, ponderile folosite sunt $\boldsymbol{f_{pl}}$ şi $\boldsymbol{x_0}$.

Pentru costurile de producţie, spre exemplu, pe baza relaţiilor generale se pot construi următorii indici ai sarcinilor de plan:

$$i^{c}_{pl/0} = \frac{c_{pl}}{c_0}; \quad i^{q}_{pl/0} = \frac{q_{pl}}{q_0} \tag{9.5}$$

şi respectiv:

$$I^{c}_{pl/0} = \frac{\sum c_{pl} q_{pl}}{\sum c_0 q_{pl}}; \quad I^{q}_{pl/0} = \frac{\sum c_0 q_{pl}}{\sum c_0 q_0} \tag{9.6}$$

unde **c** reprezintă costul pe unitatea de produs; **q** – volumul fizic al producţiei.

Din exemplele considerate se observă că indicii individuali se folosesc pentru un singur fel de produs, în timp ce în cazul producţiei eterogene, pentru care nu se poate obţine $\sum\mathbf{q}$, se utilizează indicii de grup.

9.4. Indicii agregaţi. Sisteme de ponderare folosite la construirea indicilor de grup

La nivelul unui ansamblu, valorile variabilelor statistice înregistrate pot fi însumate sau calculate sub formă de mărime medie. În primul caz (la însumare) se obţin valori agregate care trebuie, prin metoda indicilor, să fie comparate în timp şi spaţiu. Din această comparare rezultă un indice agregat. Dacă valorile individuale (parţiale) ale agregatului sunt însumabile direct nu apare nici o problemă deosebită. Dacă nu sunt însumabile, decât printr-un alt element, atunci acest etalon poartă denumirea de **pondere** şi ea trebuie să fie aleasă cu discernământ.

În practica statistică, problemele cele mai dificile apar în legătură cu alegerea şi folosirea ponderilor la construirea indicilor de grup.

Alegerea şi folosirea sistemelor de ponderare trebuie să se facă în mod diferenţiat ţinându-se seama de conţinutul indicatorului comparat, de natura datelor existente în evidenţa curentă şi posibilitatea de a stabili o analogie între descompunerea de factori a modificării absolute şi relative.

Material cu drept de autor

Pe măsura dezvoltării statisticii s-au propus mai multe sisteme de ponderare care au fost particularizate, de regulă, pe exemplul indicelui volumului fizic sau al prețurilor producției şi circulației mărfurilor. Pentru generalizare însă se vor prezenta sistemele de indici pentru o variabilă comp0lexă ($\mathbf{y_i}$), dependentă de un factor calitativ ($\mathbf{x_i}$) şi un factor cantitativ ($\mathbf{f_i}$), adică $\mathbf{y_i = x_i.f_i}$.

Având trei variabile înregistrate la nivelul unităților complexe care formează în mod permanent colectivitatea supusă observării, înseamnă că se pot calcula trei indici individuali şi trei indici de grup.

Indicii individuali se calculează ca indici simpli folosind datele înregistrate pentru fiecare variabilă la nivelul unității de observare folosită:

$$i^{y}_{1/0} = \frac{y_1}{y_0}; \quad i^{f}_{1/0} = \frac{f_1}{f_0}; \quad i^{x}_{1/0} = \frac{x_1}{x_0} \tag{9.7}$$

În acest caz, indicii de grup la nivelul întregului ansamblu se calculează ca indici agregați. Pentru prezentarea diferitelor sisteme de ponderare se presupune că variabila $\mathbf{y_i}$ este însumabilă direct şi că se descompune la nivelul fiecărei unități în produsul dintre variabila $\mathbf{x_i}$ – cu caracter de mărime statistică derivată – şi variabila $\mathbf{f_i}$ cu caracter de variabilă cantitativă neînsumabilă direct. Aceasta înseamnă că pentru indicii de grup ai celor două variabile – factori se va folosi tot un indice agregat în care succesiv factorul indexat este variabil, iar celălalt are caracter de pondere.

Rezultă, deci, că într-un sistem de indici – cel puțin la nivelul întregului ansamblu – indicii factoriali trebuie să cuprindă aceleaşi elemente ca şi indicele variabilei complexe.

În cazul variabilei complexe $\mathbf{y_i = x_i.f_i}$, indicele de grup va fi:

$$I^{y}_{1/0} = \frac{\sum_{i=1}^{k} y_{i1}}{\sum_{i=1}^{k} y_{i0}} = \frac{\sum_{i=1}^{k} x_{i1} f_{i1}}{\sum_{i=1}^{k} x_{i0} f_{i0}} = \frac{\sum x_i f_i}{\sum x_0 f_0} \tag{9.8}$$

Indicii factoriali derivați din acesta $I^{y(f)}_{1/0}$ şi $I^{y(x)}_{1/0}$ trebuie să prezinte variația unui singur factor, iar celălalt să rămână constant, deci să joace rolul de pondere şi anume:

$$I^{y(f)}_{1/0} = \frac{\sum x f_i}{\sum x f_0} \quad \text{şi} \quad I^{y(x)}_{1/0} = \frac{\sum x_i f}{\sum x_0 f} \tag{9.9}$$

La rândul lor, ponderile nu sunt elemente abstracte ci ele provin fie din perioada curentă, fie din cea de bază (în cazul indicilor de dinamică). Deci, ele trebuie supuse unei alegeri.

Material cu drept de autor

Sistemul de ponderare al lui Étienne Laspéyres (1834-1913), economist german, profesor la Universitățile din Basel, Riga şi Karlsruhe, ocupându-se de prețuri a elaborat în 1864 un indice care îi poartă numele. Astfel, el a propus un sistem de ponderare la care ponderile folosite sunt cele din perioada de bază. În acest caz, indicii factoriali se calculează pe baza relațiilor:

- pentru factorul cantitativ:

$$I_{1/0}^{y(f)} = \frac{\sum x_0 f_1}{\sum x_0 f_0} \tag{9.10}$$

- pentru factorul calitativ.

$$I_{1/0}^{y(x)} = \frac{\sum x_1 f_0}{\sum x_0 f_0} \tag{9.11}$$

Adaptarea sau respingerea celor doi indici nu se poate face decât după ce se anulează conținutul lor şi măsura în care ei reflectă nişte proporții reale cu privire la dezvoltarea fenomenelor la care se referă. Fiecare indice trebuie analizat separat, corespunzător cu conținutul indicatorilor absoluți pe care-i conține şi a relațiilor de interdependență dintre fenomenul de indexat şi ponderile folosite.

Sistemul de ponderare al lui Hermann Paasche (1851-1925), economist german, care a elaborat în 1874 un indice al prețurilor ce-i poartă numele, în care ponderile sunt cele din perioada curentă:

- pentru indicele factorului cantitativ:

$$I_{1/0}^{y(f)} = \frac{\sum x_1 f_1}{\sum x_1 f_0} \tag{9.12}$$

- pentru indicele factorului calitativ:

$$I_{1/0}^{y(x)} = \frac{\sum x_1 f_1}{\sum x_0 f_1} \tag{9.13}$$

Şi expresiile (9.12) şi (9.13) trebuie analizate în raport cu conținutul şi scopul analizei.

Sistemul de ponderare al lui Francis Edgeworth (1845-1926), economist şi matematician din şcoala anglo-saxonă, profesor la Londra şi Oxford, care şi-a adus contribuția la dezvoltarea teoriei indicilor. El a elaborat un indice pe baza unui sistem de ponderare ce ține seama de ponderile din ambele perioade. Astfel, ***indicele prețurilor calculat de Francis Edgeworth*** se bazează pe cumularea cantităților din perioada de bază cu cele din perioada curentă şi folosirea acestora ca pondere la măsurarea variației relative a prețurilor. Prețul fiind o variabilă calitativă, relația generală din care rezultă indicele este:

Material cu drept de autor

$$I_{1/0}^{(x)} = \frac{\sum x_1 (f_1 + f_0)}{\sum x_0 (f_1 + f_0)} \qquad (9.14)$$

Acest indice prezintă dezavantajul principal că poate fi particularizat numai pentru variația unui factor calitativ, iar ponderea este factorul cantitativ ale cărei valori pot fi însumate, atât de la o unitate la alta, cât şi în timp. Neputându-se extinde şi la analiza variației factorului cantitativ, acest indice nu poate fi cuprins într-un sistem de indici în cadrul căruia să se poată stabili gradul de influență a diferiților factori asupra fenomenului complex pe care-l determină.

Sistemul de ponderare al lui Irving Fischer (1867-1947), economist american, profesor la Universitatea din Yale. A contribuit la dezvoltarea teoriei indicilor, formulând testele de verificare a lor. Este autorul cunoscutului ***indice ideal*** al prețurilor care se practică de statistica multor țări. El a calculat acest indice de grup al prețurilor ca o medie geometrică a celor doi indici agregați, de tip Laspeyres şi de tip Paasche. Prin generalizare, pentru indicele variabilei calitative se obține:

$$I_{1/0}^{y(x)} = \sqrt{\frac{\sum x_1 f_0}{\sum x_0 f_0} \times \frac{\sum x_1 f_1}{\sum x_0 f_1}} \qquad (9.15)$$

Acest indice prezintă avantajul că se încadrează în intervalul de variație a valorilor indicilor calculați pe baza celor două sisteme de ponderare, deci va compensa o parte din tendința de modificare a ponderilor folosite. Practic, prezintă însă dezavantajul că necesită cunoaşterea separată a tuturor elementelor de calcul şi combinarea tuturor variantelor posibile.

Indicele lui Fischer se poate aplica pentru orice caracteristică statistică a cărei variație se măsoară cu ajutorul unui indice ponderat. Ca atare, indicele factorului cantitativ de tip Fischer se bazează pe relația:

$$I_{1/0}^{y(f)} = \sqrt{\frac{\sum x_0 f_1}{\sum x_0 f_0} \times \frac{\sum x_1 f_1}{\sum x_1 f_0}} \qquad (9.16)$$

Indicele lui Fischer se foloseşte în special în calculul indicilor teritoriali pe plan internațional.

9.5. Principii de bază ale aplicabilității indicilor agregați

La construirea indicilor de grup se manifestă particularitățile specifice ale metodei indicilor, dificultățile metodologice de calcul. De aceea se impune sistematizarea unor principii de calcul.

Aşa cum s-a observat anterior, un fenomen complex supus studiului are un dublu caracter: el apare sub forma produsului a doi indicatori din care unul este indicator de volum (cantitativ) notat $\mathbf{f_i}$, iar celălalt – indicator de calitate (calitativ) notat cu $\mathbf{x_i}$, conform convenției prezentate anterior.

Material cu drept de autor

Ambii indicatori pot fi consideraţi ca elemente ale fenomenului complex, iar acesta din urmă, ca indicator funcţional. Prin ***indicator funcţional*** se înţelege expresia unui sistem de caracteristici, determinat în mod obiectiv pe baza unei analize calitative multilaterale. În cadrul acestei analize se stabileşte şi modul de ponderare a factorilor constitutivi.

Existenţa, însă, în teoria şi practica statistică a mai multor sisteme de ponderare provine din faptul că nici una din formulele de calcul propuse nu satisface integral nici teoria şi nici practica folosirii acestei metode pentru studiul variaţiei complexe a fenomenelor. În plus, dacă în plan teoretic problema este mai uşor de rezolvat, pe plan practic apar uneori dificultăţi aproape de neînlăturat. Este suficient, de exemplu, să fie luat în discuţie indicele volumului circulaţiei mărfurilor $\mathbf{v = p.q}$, în care $\mathbf{p = x}$ reprezintă preţul – element calitativ şi $\mathbf{q = f}$ reprezintă producţia comercializată – element cantitativ şi, deci, **v** este indicator funcţional, care ar trebui să reflecte, de exemplu, numai modificarea „pură" a preţurilor.

Complexitatea unui fenomen poate fi apreciată şi dintr-un alt punct de vedere şi anume, dacă este vorba de o masă de obiecte eterogene, în care caz caracteristica lui generală este determinată de mărimea: $\sum \mathbf{y} = \sum \mathbf{f_i x_i}$. În cazul particular, când este supus studiului un singur obiect omogen, lipseşte semnul sumei: $\mathbf{y = f.x}$. În diferite probleme concrete elementul funcţional poate fi, în toate cazurile, însumat.

Pentru a se înţelege mai bine dificultăţile alegerii sistemelor de ponderare la analiza dinamicii fenomenelor economice se vor prezenta unele situaţii concrete privind opţiunea concretă pentru stabilirea indicilor agregaţi care măsoară dinamica valorii, volumul fizic şi dinamica preţurilor.

Indicele de grup al valorii se calculează ca un raport între valoarea din perioada curentă a bunului economic luat în considerare şi valoarea acestuia din perioada de bază.

Acest indice se construieşte pornind de la considerentul că valoarea (**v**) este egală cu produsul dintre cantitatea (**q**) de bunuri economice şi preţul individual al acestora (**p**). Se notează $\mathbf{v_0 = q_0 p_0}$ valoarea individuală a bunurilor economice în perioada de bază şi $\mathbf{v_1 = q_1 p_1}$ valoarea individuală a aceluiaşi bun economic în perioada curentă. Ca urmare, indicele de grup al valorii va avea forma:

$$I_{1/0}^{v(q.p)} = \frac{\sum q_1 p_1}{\sum q_0 p_0} \qquad (9.17)$$

Acest indice exprimă modificarea relativă a valorii sub influenţa celor doi factori: cantitativ (**q**) şi calitativ (**p**).

Variaţia absolută a valorii sub influenţa celor doi factori se calculează ca diferenţă dintre numărătorul şi numitorul expresiei (9.17).

$$\Delta v^{(p.q)} = \sum q_1 p_1 - \sum q_0 p_0 \qquad (9.18)$$

Indicele de grup al preţului se construieşte fie pornind de la sistemul de ponderare propus de Laspeyres, prin care variaţia

Material cu drept de autor

prețurilor se ponderează cu cantitățile din perioada de bază, fie pornind de la sistemul de ponderare al lui Paasche, prin care variația prețurilor se ponderează cu cantitățile din perioada curentă. În primul caz, se merge pe ipoteza că s-au modificat numai prețurile, iar cantitățile au rămas neschimbate, deci este vorba de a reflecta numai modificarea „pură a prețurilor". Această ipoteză nu se verifică în practică.

De aceea, în practica statistică prețul nu poate fi izolat de cantitate şi, ca urmare, producătorul, vânzătorul sau cumpărătorul resimt variația de prețuri în legătură directă cu produsul creat. Din acest motiv indicele de grup al prețului se construieşte cu cel de al doilea sistem de ponderare, folosind drept ponderi cantitățile din perioada curentă.

Deci, indicele de grup al prețurilor se calculează ca un raport între valoarea din perioada curentă şi valoarea din perioada de bază – ponderată cu cantitățile perioadei curente, adică:

$$I_{1/0}^{v(p)} = \frac{\sum q_1 p_1}{\sum q_1 p_0} \tag{9.19}$$

Acest indice exprimă cu cât s-a modificat, în mărimi relative, valoarea bunului economic sub influența variației prețurilor.

Modificarea absolută a valorii sub influența prețurilor se determină cu expresia:

$$\Delta v^{(p)} = \sum q_1 p_1 - \sum q_1 p_0 \tag{9.20}$$

Indicele de grup al volumului fizic, potrivit unei relații de sistem, se calculează folosind sistemul de ponderare Laspeyres, ca un raport între valoarea din perioada curentă, ponderată cu prețurile perioadei de bază, şi valoarea din perioada de bază, după relația:

$$I_{1/0}^{v(q)} = \frac{\sum q_1 p_0}{\sum q_0 p_0} \tag{9.21}$$

Acest indice exprimă modificarea relativă a valorii sub influența volumului fizic.

Modificarea absolută a valorii sub influența volumului fizic se calculează cu expresia:

$$\Delta v^{(q)} = \sum q_1 p_0 - \sum q_0 p_0 \tag{9.22}$$

Între cei trei indici există relația:

$$I_{1/0}^{v(p.q)} = I_{1/0}^{v(p)} \times I_{1/0}^{v(q)} \tag{9.23}$$

De asemenea, există o relație şi între modificările absolute:

$$\Delta v^{(p.q)} = \Delta v^{(p)} + \Delta v^{(q)} \tag{9.24}$$

Avantajul acestor două relații constă în faptul că este suficient ca din diferite surse de informații să se cunoască doi indici ca să se poată determina al treilea.

Material cu drept de autor

Exemplu de calcul 9.1. Pentru a se prezenta metodologia de calcul a indicilor de grup sub formă agregată, se consideră că o unitate economică produce trei produse pentru care se cunosc cantitățile produse şi prețurile unitare pentru perioada de bază şi perioada curentă, tabelul 9.2.

Tabelul 9.2

Date privind circulația mărfurilor

Produs	Unitate de măsură	Cantitatea (buc.)		Prețul unitar (u.m.)		Valoarea producției (u.m.)		
		Perioada de bază	Perioada curentă	Perioada de bază	Perioada curentă	Perioada de bază	Perioada curentă	Perioada curentă la prețurile perioadei de bază
-	-	q_0	q_1	p_0	p_1	p_0q_0	p_1q_1	p_0q_1
a	b	1	2	3	4	5	6	7
A	tone	50	52	1000	1100	50000	57200	12000
B	buc	200	220	60	63	12000	13860	13200
C	buc	500	570	100	90	50000	51300	57000
Σ	-	-	-	-	-	-	112000	122360	122200

Datele din tabelul 9.2 (col. 1-6) permit analizarea modificării relative şi absolute a cantităților şi prețurilor unitare pe fiecare produs şi pe total.

Modificarea relativă pe fiecare produs se stabileşte pe baza indicilor individuali, iar modificarea absolută se calculează ca diferență dintre nivelul din perioada curentă şi cel din perioada de bază.

Indicii individuali

$$i^q_{1/0} = \frac{q_1}{q_0} \begin{cases} i^q_A = \dfrac{52}{50} = 1{,}04 \\ i^q_B = \dfrac{220}{200} = 1{,}10 \\ i^q_C = \dfrac{570}{500} = 1{,}14 \end{cases}$$

Modificarea absolută

$$\Delta q_A = 52 - 50 = 2 \text{ tone}$$

$$\Delta q_B = 220 - 200 = 20 \text{ buc.}$$

$$\Delta q_C = 570 - 500 = 70 \text{ buc.}$$

Material cu drept de autor

$$i_{1/0}^{p}=\frac{p_1}{p_0}\begin{cases} i_A^p=\dfrac{1100}{1000}=1{,}10 & \Delta p_A=1100-1000=100\ \text{u.m.}\\ i_B^p=\dfrac{63}{60}=1{,}05 & \Delta p_B=63-60=3\ \text{u.m.}\\ i_C^p=\dfrac{90}{100}=0{,}90 & \Delta p_C=90-100=-10\ \text{u.m.}\end{cases}$$

$$i_{1/0}^{v}=\frac{q_1p_1}{q_0p_0}\begin{cases} i_A^q=\dfrac{57200}{50000}=1{,}144 & \Delta v_A=57200-50000=7200\ \text{u.m.}\\ i_B^q=\dfrac{13860}{12000}=1{,}155 & \Delta v_B=13860-12000=1860\ \text{u.m.}\\ i_C^q=\dfrac{51300}{50000}=1{,}026 & \Delta v_C=51300-50000=1300\ \text{u.m.}\end{cases}$$

Pe ansamblul celor trei produse, valoarea producției a crescut în perioada curentă față de perioada de bază de 1,0925 ori sau a reprezentat în perioada curentă 109,25% față de perioada de bază, sau a crescut cu 9,25%. Aceste concluzii rezultă din indicele de grup al valorii producției:

$$I_{1/0}^{v(q.p)}=\frac{\sum v_1}{\sum v_0}=\frac{\sum q_1p_1}{\sum q_0p_0}=\frac{122300}{112000}=1{,}0925\,sau\,109{,}25\%$$

Valoarea producției a crescut în perioada curentă cu 10360 u.m.

$$\Delta v=\sum q_1p_1-\sum q_0p_0=122360-112000=10360u.m.$$

Pornind de la faptul că **v = q.p**, înseamnă că modificarea valorii producției trebuie analizată şi explicată pornind de la schimbările intervenite în nivelul cantităților şi prețurilor unitare la fiecare produs în parte. Evidențierea influenței celor doi factori (**q** şi **p**) asupra modificării producției, pe ansamblu, presupune calcularea indicelui de grup al volumului fizic şi al indicelui prețurilor.

Indicele volumului fizic se calculează în practică ca un indice de tip Laspeyres:

$$I_{1/0}^{q}=\frac{\sum q_1p_0}{\sum q_0p_0}=\frac{122200}{112000}=1{,}091\ \textbf{sau}\ 109{,}1\%$$

Rezultă că valoarea producției, pe ansamblul celor trei produse, trebuie să crească în condițiile în care s-ar fi modificat numai producția fizică (deci prețurile ar fi rămas la nivelul anului de bază) de 1,091 ori sau în mărime absolută creşterea trebuie să fie de 10200 u.m.

$$\Delta v^q=\sum q_1p_0-\sum q_0p_0=122200-112000=10200u.m.$$

Material cu drept de autor

sau:

$$\Delta v^q = \sum (q_1 - q_0).p_0 = \sum \Delta q.p_0 = 2\times 100 + 20\times 60 + 70\times 100 = 10200 u.m.$$

Indicele preţurilor se calculează cu un indice de tip Paasche:

$$I_{1/0}^p = \frac{\sum q_1 p_1}{\sum q_1 p_0} = \frac{122360}{122200} = 1{,}001 \; sau \; 100{,}1\%$$

Deci, modificarea preţurilor unitare, în condiţiile folosirii drept pondere a cantităţilor din producţia curentă, trebuie să conducă la creşterea valorii producţiei cu 0,1%, ceea ce înseamnă că valoarea producţiei ar fi reprezentat în perioada curentă 100,1% faţă de perioada de bază.

Modificarea absolută a valorii producţiei trebuie să fie, în aceste condiţii, de +160 u.m., mărime ce rezultă din:

$$\Delta v^p = \sum q_1 p_1 - \sum q_1 p_0 = 122360 - 122200 = 160 u.m.$$

sau:

$$\Delta v^p = \sum (p_1 - p_0).q_1 = \sum \Delta p.q_1 = 100\times 52 + 3\times 220 + (-10)\times 570 = 160 u.m.$$

Pornind de la relaţia dintre cei trei indici, indicele valorii este egal cu produsul dintre indicele volumului fizic şi indicele preţului respectiv:

$$i_{1/0}^v = i_{1/0}^q \times i_{1/0}^p \quad \text{şi} \quad I_{1/0}^v = I_{1/0}^q \times I_{1/0}^p \qquad (9.25)$$

şi de la modificarea absolută a valorii producţiei, care este egală cu suma modificărilor absolute determinate de cei trei factori:

$$\Delta v = \Delta v^q + \Delta v^p \qquad (9.26)$$

contribuţia relativă a celor doi factori la modificarea valorii producţiei se prezintă astfel:

Tabelul 9.3

Determinarea contribuţiei factorului de producţie şi a factorului de preţ la valoarea producţiei

Produse	Modificare absolută			Contribuţia procentuală	
	Totală v(q,p)	din care		$\frac{\Delta v^q}{\Delta v}100$	$\frac{\Delta v^p}{\Delta v}100$
		v(q)	v(p)		
A	+7200	+2000	+5200	27,8	72,2
B	+1860	+1200	+660	64,5	35,5
C	+1300	+7000	-5700	538,5	438,5
Total	+10360	+10200	+160	98,5	1,5

Rezultă faptul că sporirea valorii producţiei a fost determinată, în principal, de creşterea producţiei fizice, factor ce deţine 98,5% din creşterea valorii producţiei.

Metodologia prezentată are şi unele neajunsuri. Astfel, dacă au apărut produse noi sau au fost eliminate unele produse din perioada curentă,

Material cu drept de autor

atunci relația de sistem $I^{v}_{1/0} = I^{q}_{1/0} \times I^{p}_{1/0}$ nu mai este satisfăcută. Pentru a lărgi cât mai mult gama de produse comparabile, în practică se pune problema să se calculeze un indice de prețuri independent, iar indicele volumului fizic să se obțină ca un raport între indicele valorii şi indicele prețurilor.

Problema includerii produselor noi în calculul indicilor se pune şi se rezolvă diferențiat în cazul indicelui volumului fizic şi al prețurilor.

La determinarea indicelui volumului fizic cantitățile din perioada curentă şi din perioada de bază sunt evaluate la aceleaşi prețuri şi anume din perioada de bază:

$$\frac{\sum q_1 p_0}{\sum q_0 p_0}$$

Produsele noi însă nu au asemenea prețuri. Necuprinderea lor, din acest motiv, iar mărimea agregatului $\sum \mathbf{q_1 p_0}$ ar însemna să se denatureze conştient concluziile privitoare la rezultatele activității unității sau ramurii în cauză. În rezolvarea acestei probleme se pot folosi mai multe căi.

O primă cale ar consta în evaluarea produselor noi la prețurile la care au apărut, deci, la prețurile efective. Aceasta ar însemna, însă, să se folosească două prețuri pentru obținerea indicatorului din numărătorul indicelui volumului fizic.

O a doua cale posibilă constă în recalcularea prețurilor produselor noi prin corectarea prețurilor efective ale acestora, pe baza unor indici de prețuri calculați pentru producția comparabilă, corespunzătoare grupei de produse din care fac parte produsele noi.

Indicele agregat armonic şi aritmetic. După cum rezultă din cele prezentate anterior, în expresia indicelui de grup al volumului fizic şi a indicelui de grup al prețului, se utilizează atât prețurile din perioada curentă cât şi prețurile din perioada de bază. De cele mai multe ori, în cadrul unor compartimente ale societăților comerciale: contabilitate, producție, desfacere etc., produsele se urmăresc la prețurile perioadei curente ($\mathbf{q_1 p_1}$).

Dar, în structura relațiilor celor doi indici, intră şi valoarea produselor calculată la prețurile perioadei de bază ($\mathbf{q_1 p_0}$). Deoarece este dificil să se țină o evidență a produselor şi la prețul perioadei de bază, atunci, pentru operativitate şi simplificare, se recurge la dinamica prețurilor pe fiecare produs, adică la indicii individuali. Raționamentul îl vom aplica separat pentru cei doi indici.

Pentru ***indicele de grup al prețului*** se porneşte de la formula propusă de Paasche şi anume:

$$I^{v(p)}_{1/0} = \frac{\sum q_1 p_1}{\sum q_1 p_0}$$

În această relație, se cunoaşte numărătorul din datele existente în evidențele curente, dar nu se cunoaşte numitorul. Se apelează la dinamica prețului (indicele individual al prețului $i^{p}_{1/0} = p_1 / p_0$). În continuare, se

Material cu drept de autor

explicitează p_0: $p_0 = p_1 / i_{1/0}^p$, se introduce la numitorul indicelui de grup şi se va obține:

$$I_{1/0}^{v(p)} = \frac{\sum q_1 p_1}{\sum \frac{1}{i_{1/0}^p} q_1 p_1} \qquad (9.27)$$

în care $\mathbf{q_1 p_1}$ este valoarea din perioada curentă, deja cunoscută în evidențele curente, jucând rol de pondere; $i_{1/0}^p$ - indicele individual al prețului, uşor de stabilit pe fiecare produs în parte.

Relația (9.27) nu este altceva decât o ***medie armonică ponderată***, iar indicele prețului se mai numeşte şi ***indicele de grup armonic*** al prețului.

El exprimă modificarea relativă a valorii produselor sub influența dinamicii individuale a prețurilor.

Modificarea absolută se calculează ca diferență dintre numărător şi numitor, cu relația:

$$\Delta v^p = \sum q_1 p_1 - \sum \frac{1}{i_{1/0}^p} q_1 p_1 \qquad (9.28)$$

Pentru calculul acestui indice se poate folosi un tabel de felul următor.

Produse	Valoarea producției în perioada curentă (q_1p_1)	Modificarea prețurilor unitare	$\frac{1}{i_{1/0}^p} q_1 p_1$
0	1	2	3
...	...	...	...
Total	$\Sigma \mathbf{q_1 p_1}$		$\sum \frac{1}{i_{1/0}^p} q_1 p_1$

În coloana 2, modificarea poate fi scrisă fie direct sub formă de indice individual ($i_{1/0}^p$) exprimat în procente sau coeficient, fie sub formă de ritm $\mathbf{R^p} = (i_{1/0}^p - 1)/100$, de unde $i_{1/0}^p = \mathbf{R^p + 100}$, exprimat în procente.

Prin raportarea coloanei 1 la 3, se obține indicele de grup armonic al prețului.

Pentru ***indicele de grup al volumului fizic*** se porneşte de la expresia propusă de Laspeyres şi anume:

$$I_{1/0}^{v(q)} = \frac{\sum q_1 p_0}{\sum q_0 p_0}$$

În această relație se cunoaşte numitorul din datele evidenței curente, în schimb nu se cunoaşte numărătorul. La fel ca în cazul precedent, se apelează la un indice individual, de data aceasta indicele individual al

Material cu drept de autor

volumului fizic, $i_{1/0}^{q} = q_1/q_0$. În continuare se explicitează $\mathbf{q_1}$: $q_1 = i_{1/0}^{q} \times q_0$ şi se introduce în relația indicelui de grup al volumului fizic şi se obține:

$$I_{1/0}^{v(q)} = \frac{\sum i_{1/0}^{q} q_0 p_0}{\sum q_0 p_0} \qquad (9.29)$$

în care $\mathbf{q_0p_0}$ reprezintă valoarea din perioada de bază existentă în evidențele curente (joacă rolul de pondere); $i_{1/0}^{q}$ este indicele individual al volumului fizic, uşor de determinat pentru fiecare produs.

Relația (9.29) este ***o medie aritmetică ponderată*** a indicilor individuali, iar indicele volumului fizic se mai numeşte ***indicele de grup aritmetic al volumului fizici***.

Acest indice exprimă modificarea relativă a produselor sub influența dinamicii individuale a volumului fizic.

Modificarea absolută se calculează ca diferență dintre numărătorul şi numitorul indicelui, cu relația:

$$\Delta v^{q} = \sum i_{1/0}^{q} q_0 p_0 - \sum q_0 p_0 \qquad (9.30)$$

Pentru calculul acestui indice se poate folosi un tabel de felul următor.

Produse	Valoarea producției în perioada curentă (q_0p_0)	Modificarea prețurilor unitare	$i_{1/0}^{q} \times q_0 p_0$
0	1	2	3
...	...	...	...
Total	$\Sigma q_0 p_0$		$\sum i_{1/0}^{q} \times q_0 p_0$

În coloana 2, modificarea volumului fizic poate fi scrisă fie direct sub formă de indice individual $i_{1/0}^{q}$ - procent sau coeficient, fie sub formă de ritm: $\mathbf{R^q = (i_{1/0}^{q} - 1)}$ de unde $i_{1/0}^{q} = \mathbf{R^q + 1}$.

Prin raportarea coloane 3 la coloana 1, se obține indicele de grup aritmetic al volumului fizic.

Indicele agregat armonic şi aritmetic se aplică şi la alte variabile din domeniul economic cum ar fi: salariul mediu, productivitatea medie, costul mediu, recolta medie, eficiența fondurilor fixe etc., adică la majoritatea mărimilor relative de intensitate.

Relațiile de calcul al indicilor de grup ca medie a indicilor individuali sunt:

$$I_{1/0}^{y(x)} = \frac{\sum x_1 f_1}{\sum \frac{1}{i_{1/0}^{x}} x_1 f_1} \quad \text{şi} \quad I_{1/0}^{y(f)} = \frac{\sum i_{1/0}^{f} x_0 f_0}{\sum x_0 f_0} \qquad (9.31)$$

Material cu drept de autor

în care: x_0f_0 şi x_1f_1 reprezintă nivelurile totalizatoare din perioada de bază şi perioada curentă.

De menționat că raționamentele de construire a celor doi indici agregați, armonic sau aritmetic, se aplică şi în cazul când se schimbă sistemul de ponderare.

9.6. Indicii nivelurilor medii

În teoria şi practica economică, se întâlnesc adesea variabile calitative care au caracter de medii. Astfel de variabile sunt: ***productivitatea medie a muncii, salariul mediu, prețul mediu pe produs, rata medie a rentabilității, eficiența medie a fondurilor fixe*** etc.

Dinamica acestor variabile medii se calculează cu un sistem special de indici care sunt denumiți ***indicii nivelurilor medii***, indici obişnuiți ca raport a două medii. Se precizează că nivelul mediu, la nivelul ansamblului, se obține ca o medie aritmetică ponderată a variabilei calitative, ponderată cu factorul cantitativ, după următoarea relație:

$$\bar{x} = \frac{\sum xf}{\sum f} \qquad (9.32)$$

în care **x** este variabila calitativă; **f** – factorul cantitativ.

Dinamica acestor niveluri medii se face utilizând sistemul de indici ai nivelurilor medii şi anume: indicele cu structură variabilă, indicele cu structură fixă şi indicii cu variații în structură.

Indicele cu structură variabilă ($I_{1/0}^{\bar{x}(x,f)}$) se calculează ca un raport între nivelul mediu din perioada curentă şi acelaşi nivel mediu din perioada de bază cu expresia:

$$I_{1/0}^{\bar{x}(x,f)} = \frac{\bar{x}_1}{\bar{x}_0} = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0} \qquad (9.33)$$

în care x_0 şi x_1 sunt valorile caracteristicii pentru care se calculează media în cele două perioade; **f** – factorul cantitativ, care joacă rol de frecvență, de pondere.

Factorul cantitativ (fiind însumabil) poate fi exprimat fie în mărimi absolute (**f**) fie în mărimi relative

$$g^f = \frac{f_i}{\sum f_i}$$

care sunt în acelaşi timp şi greutăți specifice ale factorului cantitativ **f**.

Factorul $\mathbf{g^f}$ denumit şi ***factorul structural*** (de unde denumirea de ***indice cu structură variabilă***), provoacă schimbări de formă în relația indicelui cu structură variabilă, astfel:

$$I_{1/0}^{\bar{x}(x,g^f)} = \frac{\sum x_1 g_1^f}{\sum x_0 g_0^f} \qquad (9.34)$$

Material cu drept de autor

Indicele (9.34) exprimă modificarea relativă a nivelului mediu în perioada curentă față de perioada de bază atât sub influența factorului calitativ cât şi a factorului cantitativ (structural).

Modificarea absolută, a nivelului mediu, se face prin diferența dintre numărător şi numitor, cu expresia:

$$\Delta\bar{x}^{(x,g^f)} = \sum x_1 g_1^f - \sum x_0 g_0^f \qquad (9.35)$$

Indicele cu structură fixă se calculează ca un raport între media calculată pe baza nivelului totalizator din perioada curentă şi acelaşi nivel totalizator din perioada de bază ponderat cu factorul cantitativ din perioada curentă, după relația:

$$I_{1/0}^{\bar{x}(x)} = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_1}{\sum f_1} \qquad (9.36)$$

Indicele (9.36) exprimă modificarea relativă a nivelului mediu în perioada curentă față de perioada de bază, sub influența factorului calitativ.

Dacă factorul cantitativ se exprimă sub formă de greutate specifică atunci, indicele cu structură fixă devine:

$$I_{1/0}^{\bar{x}(x)} = \frac{\sum x_1 g_1^f}{\sum x_0 g_1^f} \qquad (9.37)$$

Factorul structural rămâne constant în perioada curentă, deci se aplică sistemul de ponderare Paasche. Indicele cu structură fixă poate fi construit şi cu sistemul de ponderare Laspeyres, cu menținerea ponderilor din perioada de bază.

Modificarea absolută se calculează după relația:

$$\Delta\bar{x}^{(x)} = \sum x_1 g_1^f - \sum x_0 g_1^f \qquad (9.38)$$

Expresia (9.38) permite construirea indicelui de grup armonic care pune în evidență variația fenomenului complex sub influența dinamicii individuale a factorului calitativ, astfel:

$$I_{1/0}^{\bar{x}(x)} = \frac{\sum x_1 g_1^f}{\sum \frac{1}{i_{1/0}^x} x_1 g_1^f} \qquad (9.39)$$

Indicele cu variații de structură se calculează ca un raport între nivelul mediu din perioada curentă calculat pe baza nivelului totalizat ponderat cu factorul cantitativ din perioada de bază şi nivelul mediu din perioada de bază, astfel:

$$I_{1/0}^{\bar{x}(f)} = \frac{\sum x_0 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0} \qquad (9.40)$$

Material cu drept de autor

Indicele din expresia (9.40) exprimă variația nivelului mediu sub influența factorului cantitativ.

Dacă factorul cantitativ se exprimă sub formă de greutate specifică (structură), indicele exprimă modificarea relativă a nivelului mediu sub influența factorului structural astfel:

$$I_{1/0}^{\bar{x}(g^f)} = \frac{\sum x_0 g_1^f}{\sum x_0 g_0^f} \tag{9.41}$$

Se observă că indicele (9.42) s-a ponderat după sistemul Laspeyres, dar tot aşa de bine se putea utiliza şi sistemul Paasche.

Modificarea absolută a nivelului mediu pe seama factorului structural se obține cu expresia:

$$\Delta\bar{x}^{(g^f)} = \sum x_0 g_1^f - \sum x_0 g_0^f \tag{9.42}$$

Întrucât sistemul indicilor ca raport a două medii se construieşte prin sistemele de ponderare amintite mai sus, între cei trei indici se verifică o relație multiplicativă şi anume, indicele cu structură variabilă se descompune în produsul a celor doi indici factoriali astfel:

$$I_{1/0}^{\bar{x}(x,g^f)} = I_{1/0}^{\bar{x}(x)} \times I_{1/0}^{\bar{x}(g^f)} \tag{9.43}$$

Şi între modificările absolute există o relație aditivă:

$$\Delta\bar{x}^{(x,g^f)} = \Delta\bar{x}^{(x)} + \Delta\bar{x}^{(g^f)} \tag{9.44}$$

adică modificarea absolută a nivelului mediu este egală cu suma modificărilor datorate factorilor de influență.

Pentru aplicațiile practice ale acestor indici se prezintă următoarea machetă de tabel de calcul.

Unitate statistică	Variabila calitativă		Factorul cantitativ exprimat în mărimi absolute		Factorul cantitativ exprimat în mărimi relative		Nivelul totalizator al variabilei calitative		
	Perioada de bază	Perioada curentă	Perioada de bază	Perioada curentă	Perioada de bază	Perioada curentă	(x_0f_0)	(x_1f_1)	(x_0f_1)
	(x_0)	(x_1)	(f_0)	(f_1)	g_0^f	g_1^f			
0	1	2	3	4	5	6	7	8	9
Σ	-	-	Σf_0	Σf_1	100	100	Σx_0f_0	Σx_1f_1	Σx_0f_1

Pentru calculul indicilor nivelurilor medii, se vor utiliza datele din coloanele 7, 8 şi 9 pentru nivelul totalizator al variabilei calitative şi coloanele 3 şi 4 pentru variabila cantitativă (însumabilă).

Material cu drept de autor

Dacă dorim să scoatem în evidență factorul structural, atunci se apelează la coloanele 5 şi 6 care prezintă factorul cantitativ sub formă de greutate specifică.

9.7. Descompunerea pe factori a variației unui fenomen complex folosind metoda indicilor

Cu ajutorul metodei indicilor se poate studia variația fenomenelor economico-sociale complexe, în timp şi spațiu, sub influența factorilor care le-au generat. Este cunoscut faptul că fenomenele complexe se formează ca produs a cel puțin doi factori. Astfel, fondul de salarii este egal cu produsul dintre salariul mediu nominal şi numărul de salariați; producția este egală cu produsul dintre productivitatea medie a muncii şi numărul de salariați; costul total este egal cu costul unitar şi volumul fizic; cheltuielile totale cu materia primă consumată în producție sunt egale cu produsul dintre producția fizică, cu consumul specific şi cu prețul specific al materiilor prime, deci trei factori etc. Fenomenul complex se prezintă, deci, fie sub forma unui agregat obținut ca produs al mai multor factori, dar şi sub forma unei singure variabile sintetizată ca mărime medie.

În primul caz, variabila complexă se descompune în produsul dintre două variabile: $\Sigma \mathbf{y_i} = \Sigma \mathbf{x_i}.\mathbf{f_i}$.

În cazul când fenomenul complex se prezintă sub formă de medie, atunci, ca orice medie aritmetică ponderată poate fi analizată în funcție de valorile individuale $\mathbf{x_i}$ şi de ponderile acestora $\mathbf{f_i}$:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i}$$

Dacă se are în vedere faptul că modificarea ponderii este şi o modificare de structură, iar media este un factor de influență pentru variabila complexă **y** atunci:

$$\sum y_i = \bar{x} \times \sum f_i = \frac{\sum x_i f_i}{\sum f_i} \times \sum f_i = \sum x_i g^f \times \sum f_i \qquad (9.45)$$

Prin metoda indicilor se separă influența fiecărui factor în parte şi se calculează contribuția absolută şi relativă a acestora la modificarea fenomenului complex. Operația aceasta de separare a contribuției factorilor poartă denumirea de ***descompunere factorială***.

În teoria şi practica statistică descompunerea indicelui general în produsul indicilor factoriali se numeşte ***descompunere geometrică***, iar separarea modificării absolute totale în suma modificărilor absolute, datorate factorilor, este denumită ***descompunere analitică***. Procedeele folosite cel mai frecvent în statistică, în descompunerea variației unui fenomen complex pe factori de influență, sunt ***metoda substituirii în lanț*** şi ***metoda influențelor izolate a factorilor***, denumită şi ***metoda restului nedescompus***.

Material cu drept de autor

Metoda substituției în lanț. Aplicarea acestei metode presupune că se anihilează pe rând influența factorilor menținându-se numai variația unui singur factor. Substituind în lanț factorii, înseamnă că pentru indicii factoriali şi modificările absolute se folosesc sisteme de ponderare diferite, iar numărul lor este egal cu cel al factorilor înregistrați. Indiferent de variantele de ponderare, indicele general este egal cu produsul indicilor factoriali. Dacă, de exemplu, $\mathbf{y_i = x_i f_i}$, atunci: $I^{y}_{1/0} = I^{y(x)}_{1/0} \times I^{y(f)}_{1/0}$, respectiv: $\Delta \mathbf{y(x,f)} = \Delta \mathbf{y(x)} + \Delta \mathbf{y(f)}$.

În funcție de succesiunea substituirii factorilor, pot fi două variante. Indiferent de varianta aplicată, substituirea în lanț presupune aplicarea următoarelor reguli:

- indicele influenței primului factor, de regulă cel cantitativ, se construieşte folosind drept pondere cealaltă sau celelalte variabile la nivelul perioadei de bază;
- un factor o dată substituit rămâne drept pondere la nivelul perioadei curente pe tot parcursul descompunerii pentru ceilalți indici factoriali. Practica demonstrează că în general există un singur factor cantitativ cu care se începe analiza factorială, iar ceilalți factori sunt calitativi şi se ordonează în funcție de relațiile dintre ei.

Aceste reguli se reflectă în cazul celor două variante, reprezentate grafic în figura 9.1:

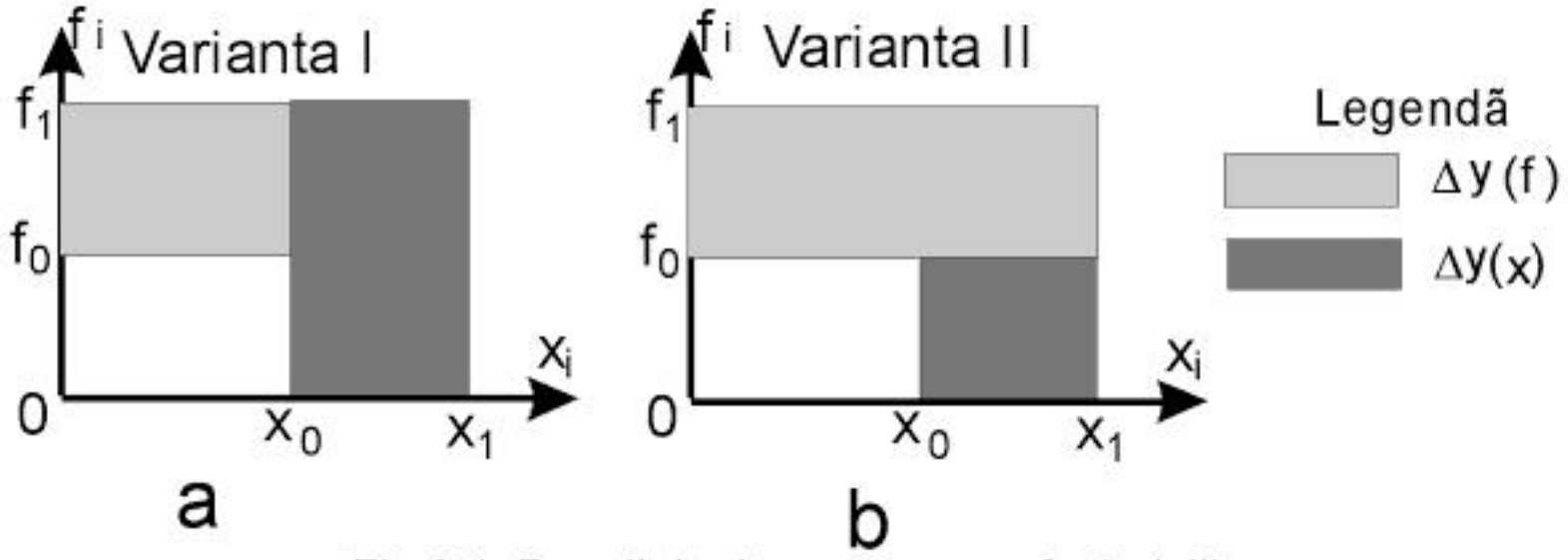


Fig.9.1. Reguli de descompunere factorială

Varianta I (figura 9.1-a): se modifică mai întâi factorul cantitativ:
a. $\mathbf{y_0 = x_0 f_0 \Rightarrow y' = x_0 f_1}$; b. $\mathbf{y' \Rightarrow y_1 = x_1 f_1}$
Varianta II (figura 9.1-b): se modifică mai întâi factorul calitativ:
a. $\mathbf{y_0 = x_0 f_0 \Rightarrow y' = x_1 f_0}$; b. $\mathbf{y' \Rightarrow y_1 = x_1 f_0}$

Indicii factoriali şi modificările absolute, corespunzătoare celor două variante, se calculează pe baza relațiilor:

Varianta I:

$$I^{y(f)}_{1/0} = \frac{\sum x_0 f_1}{\sum x_0 f_0} \qquad (9.46)$$

Material cu drept de autor

$$\Delta y(f) = \sum x_0 f_1 - \sum x_0 f_0 = \sum x_0 \Delta f \qquad (9.47)$$

$$I_{1/0}^{y(x)} = \frac{\sum x_1 f_1}{\sum x_0 f_1} \qquad (9.48)$$

$$\Delta y(x) = \sum x_1 f_1 - \sum x_0 f_1 = \sum f_1 \Delta x \qquad (9.49)$$

Varianta II:

$$I_{1/0}^{y(x)} = \frac{\sum x_1 f_0}{\sum x_0 f_0} \qquad (9.50)$$

$$\Delta y(x) = \sum x_1 f_0 - \sum x_0 f_0 = \sum f_0 \Delta x \qquad (9.51)$$

$$I_{1/0}^{y(f)} = \frac{\sum x_1 f_1}{\sum x_1 f_0} \qquad (9.52)$$

$$\Delta y(f) = \sum x_1 f_1 - \sum x_1 f_0 = \sum x_1 \Delta f \qquad (9.53)$$

Influența factorilor asupra modificării relative se calculează pe baza formulelor:

Varianta I:

Modificarea fenomenului complex **y**:

$$(I^{y(x,f)} - 1)100 = R^{y(x,f)} \qquad (9.54)$$

Modificarea pe seama factorului **f**:

$$(I^{y(f)} - 1)100 = R^{y(f)} \qquad (9.55)$$

Modificarea pe seama factorului **x**:

$$(I^{y(x)} - 1) \times I^{y(f)} 100 = R^{y(x)} \times I^{y(f)} \qquad (9.56)$$

Varianta II:

Modificarea fenomenului complex **y**:

$$(I^{y(x,f)} - 1)100 = R^{y(x,f)} \qquad (9.57)$$

Modificarea pe seama factorului **f**:

$$(I^{y(f)} - 1) \times I^{y(x)} 100 = R^{y(x)} \times I^{y(x)} \qquad (9.58)$$

Modificarea pe seama factorului **x**:

$$(I^{y(x)} - 1)100 = R^{y(x)} \qquad (9.59)$$

La construirea indicilor de grup, alegerea uneia sau alteia din cele două variante se realizează în funcție de concluziile desprinse din analiza

Material cu drept de autor

succesiunii schimbării factorilor şi de datele disponibile. În condițiile în care se cunosc valorile variantelor pentru cele două perioade se optează de regulă pentru varianta I.

Deosebirea privind mărimea cu care influențează cei doi factori modificarea variabilei complexe, în cazul celor două variante, poate fi sesizată vizual pe baza graficelor din figura 9.1, construite la nivelul unei unități de observare. Ambele variante prezentate presupun că atât variabila complexă, cât şi factorii de influență au înregistrat creşteri, respectiv $\mathbf{y_1 > y_0}$; $\mathbf{x_1 > x_0}$; $\mathbf{f_1 > f_0}$.

În condițiile în care se implică în calcul mai mult de doi factori, ordinea substituirii nu poate fi univocă, întrucât este aproape imposibil să se separeu întotdeauna riguros factorii cantitativi de cei calitativi.

Dacă modificarea variabilei complexe (**y**) se analizează în funcție de modificare a factorului ($\mathbf{x_i}$) iar factorul cantitativ se analizează nu numai ca volum ci şi ca structură, atunci se aplică o schemă trifactorială pornind de la relația: $\sum \mathbf{y} = \sum \mathbf{x^*_i f_i} \sum \mathbf{f_i}$. În acest caz se recomandă să se substituie mai întâi factorul cantitativ, urmat de factorul structural şi de cel calitativ. Indicele general:

$$\frac{\sum x_1 f_1^* \sum f_1}{\sum x_0 f_0^* \sum f_0}$$

se descompune în următorii indici factoriali:

- indicele influenței factorului cantitativ:

$$I^{y(f)} = \frac{(\sum x_0 f_0^*)\sum f_1}{(\sum x_0 f_0^*)\sum f_0} \qquad (9.60)$$

- influența factorului structural:

$$I^{y(f^*)} = \frac{(\sum x_0 f_1^*)\sum f_1}{(\sum x_0 f_0^*)\sum f_1} \qquad (9.61)$$

- influența factorului calitativ:

$$I^{y(x)} = \frac{(\sum x_1 f_1^*)\sum f_1}{(\sum x_0 f_1^*)\sum f_1} \qquad (9.62)$$

Modificarea absolută a variabilei complexe $\Delta\mathbf{y(x,f^*,\sum f)}$ se descompune în suma modificărilor datorate celor trei factori de influență:

$$\Delta y(x, f^*, \sum f) = (\sum x_0 f_0^*)\Delta\sum f + (\sum x_0 \Delta f^*)\sum f_1 + (\sum \Delta x f_1^*)\Delta\sum f_1 \qquad (9.63)$$

În practica statistică este preferată pentru comparația unor perioade scurte, varianta care presupune că se substituie mai întâi factorul cantitativ, deci folosind relațiile (9.46)...(9.49). Pentru perioade mai lungi i se impută că

Material cu drept de autor

factorii de influență sunt tratați diferențiat în sensul că acordă o importanță mai mare celui calitativ. Pentru a putea trata uniform factorii de influență implicați în analiză se recurge la metoda influențelor izolate a factorilor.

Metoda influențelor izolate a factorilor. Aplicarea acestei metode presupune că a avut loc modificarea fiecărui factor în condițiile în care ceilalți factori ar fi rămas neschimbați, deci la nivelul perioadei de bază.

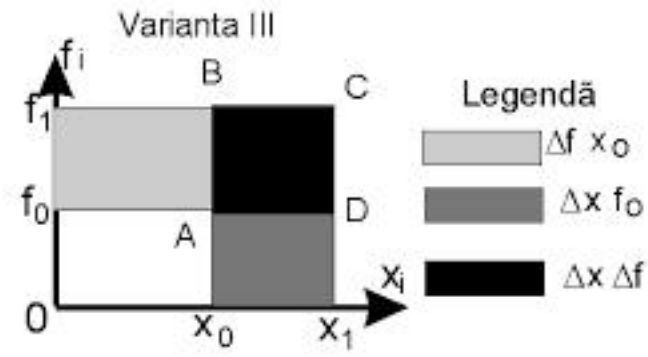


Fig.9.2. Varianta III de descompunere factorială

Pentru cazul a doi factori indicii factoriali care exprimă influența izolată a factorilor se obțin aplicând sistemul de ponderare propus de Laspeyres, şi anume:

$$I^{y(f)} = \frac{\sum x_0 f_1}{\sum x_0 f_0} \tag{9.63}$$

$$I^{y(x)} = \frac{\sum x_1 f_0}{\sum x_0 f_0} \tag{9.64}$$

Influența celor doi factori asupra modificării variabilei complexe se calculează pe baza relațiilor:

$$\Delta y(f) = \sum x_0 f_1 - \sum x_0 f_0 = \sum x_0 \Delta f \tag{9.65}$$

$$\Delta y(x) = \sum x_1 f_0 - \sum x_0 f_0 = \sum f_0 \Delta x \tag{9.66}$$

Aplicând un singur sistem de ponderare la construirea indicilor factoriali, produsul acestora nu va fi egal cu indicele variabilei complexe:

$$\frac{\sum x_0 f_1}{\sum x_0 f_0} \times \frac{\sum x_1 f_0}{\sum x_0 f_0} \neq \frac{\sum x_1 f_1}{\sum x_0 f_0}$$

iar suma modificărilor absolute datorate factorilor de influență va fi diferită de modificarea variabilei **y**:

$$(\sum x_0 f_1 - \sum x_0 f_0) - (\sum x_1 f_0 - \sum x_0 f_0) \neq (\sum x_1 f_1 - \sum x_0 f_0)$$

Mărimea cu care diferă este numită ***rest nedescompus***. Acesta apare ca urmare a faptului că indicii factoriali se construiesc folosind un singur sistem de ponderare, care nu reflectă influența variației ponderilor. Geometric, mărimea restului nedescompus poate fi sesizată cu uşurință pe baza graficului din figura 9.2, care vizualizează descompunerea pe factori a variației variabilei **y** la nivelul unității de observare.

Restul nedescompus trebuie interpretat ca fiind rezultatul influenței concomitente a celor doi factori. Cum rezultă din grafic, geometric, restul

Material cu drept de autor

nedescompus este egal cu aria ABCD, care rezultă din produsul modificării celor doi factori: $(x_1 - x_0)\times(f_1 - f_0)$.

Deci, la descompunerea variației variabilei complexe **y**, după ***metoda influențelor izolate*** este necesar să se construiască pe lângă indicii factoriali care exprimă influența izolată a factorilor **x** şi **f** şi un indice care reflectă interacțiunea celor doi factori $[I^{y(x\cap f)}]$. Corespunzător şi la descompunerea aritmetică este necesar să se calculeze modificarea variabilei **y** datorită modificării concomitente a celor doi factori $[\Delta y(x\cap f)]$.

Indicele care măsoară influența interacțiunii celor doi factori se calculează ca raport între indicele factorului calitativ construit după sistemul lui Paasche şi indicele aceleiaşi variabile construit după sistemul propus de Laspeyres:

$$I^{y(x\cap f)} = \frac{\sum x_1 f_1}{\sum x_0 f_1} : \frac{\sum x_1 f_0}{\sum x_0 f_0} \qquad (9.67)$$

Modificarea absolută a variabilei complexe rezultă din relația:

$$\Delta y(x\cap f) = (\sum x_1 f_1 - \sum x_0 f_1) - (\sum x_1 f_0 - \sum x_0 f_0) \qquad (9.68)$$

Deci, indicele variabilei complexe se descompune în produsul următorilor indici factoriali:

$$I^{y(x.f)} = \frac{\sum x_1 f_1}{\sum x_0 f_0} = \frac{\sum x_0 f_1}{\sum x_0 f_0} \times \frac{\sum x_1 f_0}{\sum x_0 f_0} \times \left[\frac{\sum x_1 f_1}{\sum x_0 f_1} : \frac{\sum x_1 f_0}{\sum x_0 f_0}\right] \qquad (9.68)$$

De aici rezultă relația pentru descompunerea analitică:

$$\begin{aligned}\sum y_1 - \sum y_0 = \sum x_1 f_1 - \sum x_0 f_0 = (\sum x_0 f_1 - \\ - \sum x_0 f_0) + (\sum x_1 f_0 - \sum x_0 f_0) + [(\sum x_1 f_1 - \\ - \sum x_0 f_1) - (\sum x_1 f_0 - \sum x_0 f_0)]\end{aligned} \qquad (9.69)$$

respectiv:

$$\Delta y(x, f) = \sum x_0 \Delta f + \sum f_0 \Delta x + \sum \Delta x.\Delta f \qquad (9.70)$$

În realitate, existând doar doi factori de influență, este obligatoriu să se separeu pe cei doi factori restul nedescompus.

Privitor la repartizarea restului nedescompus există mai multe propuneri.

a. să se atribuie integral unuia din factori, situație care conduce la procedeul substituției în lanț; factorului calitativ în varianta **I** şi factorului cantitativ în varianta **II**. Atribuirea restului nedescompus factorului calitativ, se recomandă când comparațiile se fac pe perioade scurte de timp;

b. să se repartizeze în mod egal pe factori;

c. să se repartizeze proporțional cu influențele independente ale factorilor, şi anume $\Delta f x_0$ şi $\Delta x f_0$.

Material cu drept de autor

În cea de a treia ipoteză aplicarea procedeului influențelor izolate în descompunerea pe factori se realizează în două faze:

- în prima fază se calculează influența izolată a fiecărui factor, folosind indici factoriali cu ponderi din perioada de bază, indici Laspeyres pentru ambii factori plus restul nedescompus;
- în cea de a doua fază se calculează cota parte care revine fiecărui factor din restul nedescompus (**k^x** şi **k^f**), ca un raport dintre influența independentă a fiecărui factor şi suma celor două influențe absolute independente:

$$k^x = \frac{\Delta x\, f_0}{\Delta x\, f_0 + \Delta f\, x_0} \qquad (9.71)$$

$$k^f = \frac{\Delta f\, x_0}{\Delta f\, x_0 + \Delta x\, f_0} \qquad (9.72)$$

Sporul total al variabilei $\mathbf{y_i}$ care revinde factorului **f**:

$$\Delta y(f) = \Delta f\, x_0 + k^f \frac{\Delta f\, x_0}{\Delta f\, x_0 + \Delta x\, f_0} \qquad (9.73)$$

Sporul total al variabilei $\mathbf{y_i}$ care revinde factorului **x**:

$$\Delta y(x) = \Delta x\, f_0 + k^x \frac{\Delta x\, f_0}{\Delta x\, f_0 + \Delta f\, x_0} \qquad (9.74)$$

Pornind de la influența absolută a fiecărui factor asupra modificării variabilei complexe se calculează ponderea, contribuția factorilor la formarea sporului total, respectiv:

$$\frac{\Delta y(f)}{\Delta y(f,x)}100 \text{ şi } \frac{\Delta y(x)}{\Delta y(f,x)}100 \qquad (9.75)$$

Comparativ cu substituția în lanț, metoda influențelor izolate permite explicarea mai veridică a cauzelor care au condiționat variația variabilei complexe.

Folosirea metodei influențelor izolate întâmpină însă dificultăți în condițiile în care creşte numărul factorilor de influență. Aceasta deoarece se amplifică numărul sporurilor care se datorează interacțiunii factorilor şi, o dată cu aceasta, sporeşte caracterul convențional privind atribuirea restului nedescompus al factorilor de influență. Pentru exemplificare se prezintă variația variabilei $\mathbf{y_i}$ în funcție de trei factori, pornind de la relația:

$$\sum y = \sum x.f * \sum f$$

Cei trei indici factoriali şi modificările absolute ce reflectă influența independentă sunt:

Material cu drept de autor

$$I^{y(\sum f)} = \frac{\sum x_0 f_0^* \sum f_1}{\sum x_0 f_0^* \sum f_0}; \Delta y(\sum f) = \sum x_0 f_0^* \Delta \sum f \quad (9.76)$$

$$I^{y(f^*)} = \frac{\sum x_0 f_1^* \sum f_0}{\sum x_0 f_0^* \sum f_0}; \Delta y(f^*) = \sum x_0 f^* \Delta \sum f_0 \quad (9.77)$$

$$I^{y(x)} = \frac{\sum x_1 f_0^* \sum f_0}{\sum x_0 f_0^* \sum f_0}; \Delta y(\sum x) = \sum x f_0^* \Delta \sum f_0 \quad (9.78)$$

Mărimea restului nedescompus rezultă, în acest caz, din însumarea următoarelor componente: $\Delta\sum$**f.**Δ**f***; $\Delta\sum$**f.**Δ**x**; Δ**f***Δ**x**; şi Δ**x**Δ**f***$\Delta\sum$**f**. Atribuirea celor patru sporuri nedescompuse pe cei trei factori de influență poate fi efectuată folosind una din propunerile menționate mai sus.

Exemplu de calcul 9.2. Pentru a se exemplifica aplicare acelor două metode de măsurare a influenței factorilor asupra variației unui fenomen complex, se propune să se calculeze influența productivității muncii şi a numărului salariaților asupra variației producției marfă în cadrul unei colectivități formată din două unități productive, tabelul 9.4.

Tabelul 9.4

Date de calcul pentru descompunerea factorială a unui fenomen complex

Unități de producție	Producția marfă (unități monetare		Numărul salariaților (persoane fizice)		Productivitatea muncii (u.m./persoană)		$W_0.T_1$ (u.m.)	$W_0.T_0$ (u.m.)
	Perioada de bază	Perioada curentă	Perioada de bază	Perioada curentă	$\bar{W}_0$	$\bar{W}_1$		
A	1280	1370,88	4000	4080	320,00	336,00	1305,60	1344
B	500	353,30	2000	2040	250,00	255,00	515,00	510
Σ	1780	1906,18	6000	6120	299,67	311,46	1820,60	1854
	ΣPM_0	ΣPM_1	ΣT_0	ΣT_1	ΣW_0	ΣW_1	$\Sigma W_0 T_1$	$\Sigma W_1 T_0$

Variabila complexă a cărei variație se analizează este producția marfă (**PM**), iar cei doi factori de influență: numărul salariaților (factor cantitativ, **T**) şi productivitatea muncii (factor calitativ, **W**).

Producția marfă a crescut de 1,071 ori, respectiv cu 126,18 u.m., iar numărul angajaților a sporit cu 2%, respectiv cu 120 persoane.

Material cu drept de autor

Productivitatea muncii pe fiecare unitate de producție s-a calculat ca un raport între producția marfă şi numărul personalului:

$$W_i = \frac{PM_i}{T_i}$$

iar pe total, împărțind nivelul totalizator corespunzător celor doi indicatori:

$$\overline{W} = \frac{\sum PM_i}{\sum T_i}$$

Rezultatele se prezintă în coloanele 5 şi 6 din tabelul 9.4.

Producția marfă la nivelul fiecărei unități de producție poate fi scrisă sub forma: $\mathbf{PM_i = W_iT_i}$, iar pe total $\Sigma\mathbf{PM_i = \Sigma W_iT_i}$. Deci, indicele producției marfă se prezintă sub forma unui indice agregat:

$$I^{PM} = \frac{\sum PM_1}{\sum PM_0} = \frac{\sum W_1T_1}{\sum W_0T_0} = 1{,}071 \Rightarrow \Delta\sum PM = 126{,}18\ u.m.$$

Metoda substituției în lanț.

- influența factorului cantitativ:

$$I^{\sum PM(T)} = \frac{\sum W_0T_1}{\sum W_0T_0} = \frac{1820{,}6}{1780{,}0} = 1{,}023\ sau\ 102{,}3\%$$

$$\Delta\sum PM(T) = \sum W_0T_1 - \sum W_0T_0 = 1820{,}6 - 1780{,}0 = 40{,}6\ u.m.$$

- influența factorului calitativ:

$$I^{\sum PM(W)} = \frac{\sum W_1T_1}{\sum W_0T_1} = \frac{1906{,}18}{1820{,}60} = 1{,}0473\ sau\ 104{,}7\%$$

$$\Delta\sum PM(W) = \sum W_1T_1 - \sum W_0T_1 = 1906{,}18 - 1820{,}60 = 85{,}58\ u.m.$$

$$I^{\sum PM(TW)} = I^{\sum PM(T)} \times I^{\sum PM(W)} = 1{,}023 \times 1{,}047 = 1{,}071 sau 107{,}1\%$$

$$\Delta\sum PM(TW) = \Delta\sum PM(T) + \Delta\sum PM(W) = 40{,}6 + 85{,}58 = 126{,}18\ u.m.$$

Sporul producției marfă s-a datorat în proporție de 68,82% creşterii productivității muncii şi în proporție de 31,18% creşterii numărului salariaților.

Metoda influențelor izolate.

Etapa I

- influența izolată a factorului cantitativ:

$$I^{\sum PM(T)} = \frac{\sum W_0T_1}{\sum W_0T_0} = \frac{1820{,}6}{1780{,}0} = 1{,}023\ sau\ 102{,}3\%$$

$$\Delta\sum PM(T) = \sum W_0T_1 - \sum W_0T_0 = 1820{,}6 - 1780{,}0 = 40{,}6\ u.m.$$

- influența izolată a factorului calitativ:

$$I^{\sum PM(W)} = \frac{\sum W_1T_0}{\sum W_0T_0} = \frac{1854}{1780} = 1{,}0415\ sau\ 104{,}15\%$$

$$\Delta\sum PM(W) = \sum W_1T_0 - \sum W_0T_0 = 1854 - 1780 = 74{,}0\ u.m.$$

Material cu drept de autor

- influența modificării concomitente a productivității muncii şi a numărului salariaților:

$$I^{\sum PM(W\cap T)} = \frac{\sum W_1T_1}{\sum W_0T_1} : \frac{\sum W_1T_0}{\sum W_0T_0} = \frac{1906,18}{1820,60} : \frac{1854}{1780} = \frac{1,0470}{1,0415} =$$

$$= 1,0053 \;\; sau \;\; 100,53\%$$

$$\Delta\sum PM(W \cap T) = (1906,18 - 1820,60) - (1854 - 1780) =$$

$$= 85,58 - 74,00 = 11,58 \; u.m.$$

Deci, sporul producției marfă care s-a repartizat direct, în prima etapă, pe cei doi factori a fost de 114,6 u.m. (sporul descompus), iar restul nedescompus s-a ridicat al 11,58 u.m.

$$\Delta\sum PM(W) + \Delta\sum PM(T) = 74,0 + 40,6 + 114,60 \; u.m.$$

$$\Delta\sum PM(W \cap T) = 11,58 \; u.m.$$

$$\Delta\sum PM(W) + \Delta\sum PM(T) + \Delta\sum PM(W \cap T) = 126,18 \; u.m.$$

Etapa II-a – repartizarea restului nedescompus.

- cota parte din restul nedescompus care revine factorului cantitativ:

$$K^{q(T)} = \frac{\sum \Delta TW_0}{\sum \Delta WT_0 + \sum \Delta TW_0} = \frac{74,0}{74,0 + 40,6} = 0,646 \; sau \; 64,6\%$$

- influența totală a factorului cantitativ asupra modificării producției marfă:

$$\sum \Delta TW_0 + K^{q(T)}.\Delta T.\Delta W = 40,6 + 0,354 \times 11,58 = 44,70 u.m.$$

- influența totală a factorului calitativ asupra sporului producției marfă:

$$\sum \Delta TW_0 + K^{q(W)}.\Delta T.\Delta W = 74,0 + 0,646 \times 11,58 = 81,48 u.m.$$

Sporul producției marfă de 126,18 u.m. se explică în proporție de 35,4% pe seama creşterii numărului de salariați şi în proporție de 64,6% pe seama creşterii productivității muncii:

- contribuția numărului salariaților la creşterea producției marfă:

$$\frac{\sum \Delta TW_0 + \sum K^T.\Delta W.\Delta T}{\Delta\sum PM(W.T)} \times 100 = \frac{44,70}{126,18} \times 100 = 35,4\%$$

- contribuția productivității muncii la creşterea producției marfă:

$$\frac{\sum \Delta WT_0 + \sum K^W.\Delta W.\Delta T}{\Delta\sum PM(W.T)} \times 100 = \frac{81,48}{126,18} \times 100 = 64,6\%$$

Material cu drept de autor

9.8. Serii cronologice de indici statistici

Caracterizarea evoluției fenomenelor pe perioada expirată şi fundamentarea statistică a nivelurilor dinamicii pentru perioada următoare presupun analiza seriilor cronologice. Asemenea serii se construiesc nu numai pentru indicatorii absoluți ci şi pentru cei relativi şi îndeosebi pentru indici.

În practica statistică, se construiesc ***serii de indici de grup cu bază fixă şi ponderi constante sau variabile*** şi ***serii de indici de grup cu bază în lanț şi ponderi constante sau variabile***.

Notându-se cu $\mathbf{x_i}$ seria valorilor caracteristicii şi cu $\mathbf{f_i}$ ponderile folosite, $i=\overline{1,n}$, se exemplifică tipurile de indici de grup care se pot construi în funcție de baza de comparație şi de ponderile folosite.

Indicii de grup cu bază fixă şi ponderi constante, luate din perioada de bază.

În cazul indicilor construiți pentru variabila $\mathbf{x_i}$, seria rezultă din dezvoltarea relației:

$$I_{1/0}^{x}=\frac{\sum x_i f_0}{\sum x_0 f_0}\ (i=\overline{1,n})$$

şi anume:

$$I_{1/0}^{x}=\frac{\sum x_1 f_0}{\sum x_0 f_0};I_{2/0}^{x}=\frac{\sum x_2 f_0}{\sum x_0 f_0};I_{3/0}^{x}=\frac{\sum x_3 f_0}{\sum x_0 f_0};\ldots;I_{n/0}^{x}=\frac{\sum x_n f_0}{\sum x_0 f_0};\quad (9.79)$$

Similar se obține seria de indici pentru variabila $\mathbf{f_i}$, pornindu-se de la relația generală:

$$I_{1/0}^{f}=\frac{\sum x_0 f_i}{\sum x_0 f_0}\ (i=\overline{1,n})$$

$$I_{1/0}^{f}=\frac{\sum x_0 f_1}{\sum x_0 f_0};I_{2/0}^{f}=\frac{\sum x_0 f_2}{\sum x_0 f_0};I_{3/0}^{f}=\frac{\sum x_0 f_3}{\sum x_0 f_0};\ldots;I_{n/0}^{f}=\frac{\sum x_0 f_n}{\sum x_0 f_0};\quad (9.80)$$

Indicii de grup cu bază în lanț cu ponderi constante se construiesc în felul următor.

Pentru variabila $\mathbf{x_i}$ seria de indici rezultă din relațiile:

$$I_{i/i-1}^{x}=\frac{\sum x_i f_0}{\sum x_{i-1} f_0}$$

dacă ponderea este luată din perioada de bază;

$$I_{i/i-1}^{x}=\frac{\sum x_i f_n}{\sum x_{i-1} f_n}$$

dacă ponderea este luată pentru perioada curentă.

În ambele cazuri se verifică relația care există între indicii cu baza în lanț şi cei cu baza fixă şi anume:

Material cu drept de autor

$$\frac{\sum x_1 f_0}{\sum x_0 f_0} \times \frac{\sum x_2 f_0}{\sum x_1 f_0} \times \frac{\sum x_3 f_0}{\sum x_2 f_0} \times \ldots \times \frac{\sum x_n f_0}{\sum x_{n-1} f_0} = \frac{\sum x_n f_0}{\sum x_0 f_0} \qquad (9.81)$$

respectiv:

$$\frac{\sum x_1 f_n}{\sum x_0 f_n} \times \frac{\sum x_2 f_n}{\sum x_1 f_n} \times \frac{\sum x_3 f_n}{\sum x_2 f_n} \times \ldots \times \frac{\sum x_n f_n}{\sum x_{n-1} f_n} = \frac{\sum x_n f_n}{\sum x_0 f_n} \qquad (9.82)$$

Relația (9.81) se aplică frecvent în practica statistică la calcularea şi alcătuirea seriilor de indici ai prețurilor.

Pentru variabila $\mathbf{f_i}$ seria se obține dezvoltând expresiile:

$$I^f_{i/i-1} = \frac{\sum x_0 f_i}{\sum x_0 f_{i-1}}$$

dacă ponderea este luată din perioada de bază;

$$I^f_{i/i-1} = \frac{\sum x_n f_i}{\sum x_n f_{i-1}}$$

dacă ponderea este luată pentru perioada curentă.

Şi în aceste cazuri se verifică relația dintre indicii cu bază fixă şi cel cu bază în lanț şi anume:

$$\frac{\sum x_0 f_1}{\sum x_0 f_0} \times \frac{\sum x_0 f_2}{\sum x_0 f_1} \times \frac{\sum x_0 f_3}{\sum x_0 f_2} \times \ldots \times \frac{\sum x_0 f_n}{\sum x_0 f_{n-1}} = \frac{\sum x_0 f_n}{\sum x_0 f_0} \qquad (9.83)$$

$$\frac{\sum x_n f_1}{\sum x_n f_0} \times \frac{\sum x_n f_2}{\sum x_n f_1} \times \frac{\sum x_n f_3}{\sum x_n f_2} \times \ldots \times \frac{\sum x_n f_n}{\sum x_n f_{n-1}} = \frac{\sum x_n f_n}{\sum x_n f_0} \qquad (9.84)$$

Relația (9.83) se utilizează frecvent la construirea seriilor de indici ai volumului fizic.

Indicii cu bază în lanț şi ponderi variabile. Pentru ambele variabile $\mathbf{x_i}$ şi $\mathbf{f_i}$ se pot construi indici cu ponderi variabile luate din perioada de bază şi din perioada curentă. Relațiile generale din care rezultă seria de indici sunt:

- dacă se foloseşte ponderea din perioada de bază:

$$I^x_{i/i-1} = \frac{\sum x_i f_{i-1}}{\sum x_{i-1} f_{i-1}} \qquad (9.85)$$

$$I^f_{i/i-1} = \frac{\sum x_{i-1} f_i}{\sum x_{i-1} f_{i-1}} \qquad (9.86)$$

- dacă ponderea este luată din perioada curentă;

$$I^x_{i/i-1} = \frac{\sum x_i f_i}{\sum x_{i-1} f_i} \qquad (9.87)$$

$$I^f_{i/i-1} = \frac{\sum x_i f_i}{\sum x_i f_{i-1}} \qquad (9.88)$$

Material cu drept de autor

În cazul seriilor de indici din dezvoltarea relațiilor (9.85)...(9.88), produsul indicilor cu bază în lanț şi cu ponderi variabile nu este egal cu indicele cu bază fixă al întregii perioade.

9.9. Teste de verificare a indicilor

Reversibilitatea în timp. Potrivit acestui test, indicele anului **b** calculat cu baza în anul **a**, reprezintă o mărime inversă a indicelui anului **a** calculat cu baza în anul **b**:

$$i_{b/a} = \frac{1}{i_{a/b}} \tag{9.89}$$

Reversibilitatea în timp este o proprietate pe care o au toți indicii statistici. În acest caz, comparația între două perioade nu mai depinde de perioada aleasă de bază, iar indicii ar respecta relația:

$$i_{b/a} \times i_{a/b} = 1 \tag{9.90}$$

Circularitatea. Dacă se consideră anii **a, b** şi **c** şi se calculează un indice al anului **b** cu baza în anul **a** şi un alt indice pentru anul **c** cu baza în anul **b**, produsul lor trebuie să fie egal cu indicele anului **c** cu baza în anul **a**. De remarcat că atunci când indicii răspund testului circularității, ultimul indice, care de fapt este un indice al dinamicii cu baza fixă, se poate obține direct, fără intermediul anului **b**, adică:

$$i_{b/a} \times i_{c/b} = i_{c/a} \tag{9.91}$$

şi în acest caz:

$$i_{b/a} \times i_{a/b} \times i_{a/c} = 1 \tag{9.92}$$

Circularitatea este de fapt o extindere a testului de reversibilitate.

Reversibilitatea factorilor. După această regulă, dacă se substituie factorii indicelui, produsul noilor indici nu se modifică. Acestui test îi răspund, de regulă, sistemele de indici concepute pentru analiza variației fenomenelor complexe. Dacă factorii **x** şi **f** îşi schimbă locurile, atunci rezultă că:

$$I_{1/0}^{y(x)} = \frac{\sum x_1 f_1}{\sum x_0 f_1} \Rightarrow I_{1/0}^{y(f)} = \frac{\sum f_1 x_1}{\sum f_0 x_1} \tag{9.93}$$

şi respectiv:

$$I_{1/0}^{y(f)} = \frac{\sum x_0 f_1}{\sum x_0 f_0} \Rightarrow I_{1/0}^{y(x)} = \frac{\sum f_0 x_1}{\sum f_0 x_0} \tag{9.94}$$

Produsul noilor indici factoriali obținuți este, de asemenea, egal cu indicele general al variației fenomenului complex.

Material cu drept de autor

9.10. Indici teritoriali

Indicii teritoriali au o largă aplicabilitate în compararea unor caracteristici statistice situate în unități de spațiu diferite. Un indice teritorial se calculează ca raport între termenii unei serii statistice de spațiu şi exprimă variația nivelului unor caracteristici în raport cu spațiul.

Ca la orice indice, dar mai ales la cei teritoriali, se pune problema alegerii bazei de raportare şi a sistemului de ponderare.

Alegerea bazei de raportare se face în primul rând pornind de la raționamente economice sau sociale.

La stabilirea sistemului de ponderare se aplică aceleaşi principii generale prezentate anterior.

Şi indicii teritoriali se alcătuiesc ca indici individuali (**i**) şi indici de grup (**I**).

Dacă se notează variabila ce se compară cu **y** şi cele două unități de spațiu cu **A** şi **B**, se vor obține doi indici diferiți din punct de vedere al sensului comparării, astfel:

$$i^{y}_{A/B} = \frac{y_A}{y_B} \text{ şi } i^{y}_{B/A} = \frac{y_B}{y_A} \qquad (9.95)$$

Ca şi la indicii dinamici, între cei doi indici individuali teritoriali există o relație de reversibilitate în spațiu:

$$i^{y}_{A/B} \times i^{y}_{B/A} = 1 \qquad (9.96)$$

Se face precizarea că în analizele statistice, cele două sensuri de comparație, nu se utilizează simultan.

Indicii de grup teritoriali se alcătuiesc prin raportarea nivelului fenomenului complex din spațiul **A**, la nivelul fenomenului complex din spațiul **B**. fenomenul complex se notează cu $\sum \mathbf{y_i}$, în care apare şi factorul de ponderare, frecvența. Relația generală de calcul a unui indice de grup teritorial este, în funcție de comparație:

$$I^{y}_{A/B} = \frac{\sum y_A}{\sum y_B} \quad sau \quad I^{y}_{B/A} = \frac{\sum y_B}{\sum y_A} \qquad (9.97)$$

în care: $\sum \mathbf{y_A}$ – nivelul totalizator dintr-o unitate de spațiu **A**; $\sum \mathbf{y_B}$ – nivelul totalizator al fenomenului dintr-o unitate de spațiu **B**.

Dacă ***factorul cantitativ*** este direct însumabil, atunci ***factorul calitativ*** se manifestă la nivelul ansamblului de elemente cercetate ca o medie. Este vorba, aşa cum s-a văzut anterior, de acele variabile neînsumabile direct exprimate sub formă de mărimi relative de intensitate.

Indicele de grup teritorial, în cazul acesta apare ca un raport între două medii, astfel:

Material cu drept de autor

$$I_{A/B}^{\bar{x}} = \frac{\bar{x}_A}{\bar{x}_B} = \frac{\sum y_A}{\sum f_A} : \frac{\sum y_B}{\sum f_B} = I_{A/B}^{y} \times I_{A/B}^{f} \qquad (9.98)$$

Rezultă că şi în cazul indicilor teritoriali se manifestă legătura dintre variabila complexă şi factorii săi de influență, deci:

$$I_{A/B}^{y(\bar{x},f)} = I_{A/B}^{\bar{x}} \times I_{A/B}^{f} \qquad (9.99)$$

Dacă factorul cantitativ nu este însumabil direct, atunci pentru comparațiile în spațiu se alege ponderea corespunzătoare. De exemplu, factorul cantitativ poate fi ponderat cu $\mathbf{x_A}$ sau cu $\mathbf{x_B}$ şi atunci indicele de grup care exprimă variația factorului cantitativ în spații diferite este:

$$I_{A/B}^{f} = \frac{\sum x_A f_A}{\sum x_A f_B} \text{ şi } I_{A/B}^{f} = \frac{\sum x_B f_A}{\sum x_B f_B} \qquad (9.100)$$

În cazul factorului calitativ neînsumabil, ponderea se face cu $\mathbf{f_A}$ sau cu $\mathbf{f_B}$ şi atunci indicele de grup care exprimă variația factorului calitativ în spații diferite va fi:

$$I_{A/B}^{x} = \frac{\sum x_A f_A}{\sum x_B f_A} \text{ şi } I_{A/B}^{x} = \frac{\sum x_A f_B}{\sum x_B f_B} \qquad (9.101)$$

Se observă că indicii factoriali sunt ponderați cu frecvențe specifice unităților teritoriale. În aceste condiții nu se asigură reversibilitatea factorilor. Pentru a se stabili o relație între fenomenul complex şi factorii corespunzători se foloseşte un fel de sistem de ponderare propus de Fischer, care are la bază media geometrică a celor două variabile de ponderare. Deci indicii de grup sunt:

- pentru factorul cantitativ:

$$I_{A/B}^{f} = \sqrt{\frac{\sum x_A f_A}{\sum x_A f_B} \times \frac{\sum x_B f_A}{\sum x_B f_B}} \qquad (9.102)$$

- pentru factorul calitativ:

$$I_{A/B}^{f} = \sqrt{\frac{\sum x_A f_A}{\sum x_B f_A} \times \frac{\sum x_A f_B}{\sum x_B f_B}} \qquad (9.103)$$

Atât indicii ca raport a două medii, cât şi indicii teritoriali agregați, fac obiectul cercetărilor în profil de spațiu, dar mai ales în comparațiile internaționale.

9.11. Metode de ierarhizare a unităților spațiale

Ierarhizarea unităților teritoriale, după anumite criterii, prezintă o importanță deosebită, atât în plan național, pentru dezvoltarea armonioasă a

Material cu drept de autor

tuturor unităților de spațiu, cât şi internațional, pentru a măsura decalajele dintre țări în vederea elaborării unor strategii optime de dezvoltare şi de întrajutorare.

Pentru o clasificare şi o ierarhizare judicioasă este contraindicată folosirea unui singur indicator statistic, oricât de relevant ar fi el, şi se recomandă o analiză multicriterială, prin combinarea unui sistem complet de indicatori statistici.

Orice ierarhizare începe cu identificarea şi selecționarea acelor indicatori care asigură caracterizarea multilaterală a fiecărei unități administrativ-teritoriale. După ce se fac o serie de clasamente provizorii pe baza fiecărui indicator selectat, se trece la alegerea metodei de agregare într-un singur indicator, pe baza căruia se realizează propriu-zis ierarhizarea.

Dintre metodele mai importante pot fi amintite: metoda rangurilor, metoda matricială, metoda observării distanței relative.

Metoda rangurilor se bazează pe o ierarhizare în funcție de mărimea nivelului caracteristicii, atribuindu-se ranguri (numere de ordine) fiecărei unități teritoriale. Unitatea cu un nivel al caracteristicii cel mai mare primeşte rangul unu şi aşa mai departe, în mod succesiv, până la unitatea care înregistrează nivelul cel mai mic al caracteristicii, care primeşte rangul cel mai mare.

Dacă o unitate de spațiu înregistrează mai multe caracteristici (criterii) după care se face ierarhizarea, pentru fiecare din aceste variabile se atribuie ranguri în mod succesiv. Macheta din tabelul 9.5 prezintă sintetizat metoda rangurilor.

Tabelul 9.5

Machetă de tabel pentru ierarhizarea unităților spațiale

Unitatea spațială	Rangul atribuit în funcție de				Scor total	Rang final
	Caracte-ristica A	Caracte-ristica B	Caracte-ristica C	Caracte-ristica D		
0	**1**	**2**	**3**	**4**	**5**	**6**
a	**2**	**1**	**1**	**4**	**8**	**2**
b	**1**	**3**	**2**	**1**	**7**	**1**
c	**4**	**2**	**3**	**3**	**12**	**3**
d	**3**	**4**	**4**	**2**	**13**	**4**

În coloanele 1, 2, 3, 4 sunt atribuite rangurile pentru toate cele patru caracteristici ce definesc unitatea teritorială.

La fiecare unitate teritorială se înseamnă rangurile (pe orizontală) şi se obține „scorul total" din coloana 5. În continuare, pentru scorul cel mai mic se acordă rangul 1, etc., rezultând în coloana 6 „rangul" final pe baza căruia unitățile de spațiu sunt ierarhizate după cele patru criterii (caracteristici).

Material cu drept de autor

Această metodă, a rangurilor, prezintă avantajul că este simplă şi, deci, uşor de aplicat, rezultatele ei putând fi valorificate în analizele care privesc corelațiile dintre variabile prin metode neparametrice. Prezintă însp neajunsul că prin această metodă se pierd informații cu ocazia celor două nivelări ale diferitelor valori ale caracteristicii: odată când se atribuie rangurile pentru fiecare caracteristică şi altădată când se atribuie rangurile pentru scorurile totale. De aceea se apelează la alte metode de ierarhizare.

Metoda observării distanței relative față de performanța maximă este o metodă simplă care permite păstrarea integrală a informațiilor cu privire la distanțele reale dintre nivelurile caracteristicilor înregistrate în unitățile de spațiu.

Exemplu de calcul 9.3. Vom considera, spre exemplu, ierarhizarea a unor unități teritoriale de tip service auto, luându-se în considerare beneficiul total anual în unități monetare, beneficiul specific în unități monetare pe salariat şi indicele de realizare a beneficiului în momentul 2 față de momentul 1, tabelul 9.6.

Tabelul 9.6

Date pentru exemplul de ierarhizare multicriterială a unităților spațiale

Unitatea	Beneficiul global anual u.m.	Beneficiul specific, u.m./salariat	Indicele (%) între momentele de timp: 1 şi 2.
0	1	2	3
A	2340	296	96,30
B	4500	865	86,09
C	2320	455	76,82
D	23075	402	88,24
E	27220	340	108,12
F	600	58	133,04
G	5360	1489	100,41
H	10140	170	123,16
I	14000	243	119,91
J	10800	710	83,63
K	1580	149	166,32
L	5380	136	116,85
M	3270	376	88,86
N	3330	320	94,95

În prima coloană (1) din tabelul 9.6 este prezentată o serie teritorială formată din indicatori absoluți – dimensiunile beneficiului realizat de un număr oarecare de unități de tip service. Astfel de indicatori permit o analiză formală a variației spațiale a unui indicator economic (în cazul de față dimensiunile beneficiului), rezultatele fiind puțin interesante pentru concluzii ştiințifice adecvate, întrucât nu ia în considerare domeniul de activitate de tip service, dimensiunile unității de tip service etc.

Material cu drept de autor

Comparațiile pot deveni utile şi pertinente dacă se iau în considerare indicatorii derivați din coloanele 1 şi 2, din tabelul 9.6. Fiecare nouă serie teritorială creată din indicatori derivați va aduce un aport substanțial la evaluarea caracteristicii economice analizată.

Astfel, se iau în considerare coloanele 2 şi 3 din tabelul 9.6 şi pornindu-se de la valorile maxime spre cele minime, pe coloane, se atribuie numere de la 1 la 14 ca ranguri de ordonare valorică, obținându-se coloanele 3 şi 4 din tabelul 9.7. Prin însumarea rangurilor pentru fiecare unitate de producție, de pe coloanele 3 şi 4, se completează scorul din coloana 5, din tabelul 9.7. Unitatea economică cu scorul cel mai mic este cea mai performantă din ambele puncte de vedere luate în considerare şi obține astfel rangul 1. Pe măsură ce scorul creşte, se măreşte şi rangul final, până se ajunge la rangul **n** atribuit unității teritoriale care însumează punctajul (scorul) maxim.

Tabelul 9.7

Prima prelucrare a datelor din tabelul 9.6

Unitatea	Beneficiul specific, u.m./salariat	Indicele (%) între momentele de timp: 1 şi 2.	Rangul atribuit		Scor	Rang final
			După col.1	După col.2		
0	1	2	3	4	5	6
A	296	96,30	9	8	17	11-12
B	865	86,09	2	12	14	4-6
C	455	76,82	4	14	18	13-14
D	402	88,24	5	11	16	7-10
E	340	108,12	7	6	13	2-3
F	58	133,04	14	2	16	7-10
G	1489	100,41	1	7	8	1
H	170	123,16	11	3	14	4-6
I	243	119,91	10	4	14	4-6
J	710	83,63	3	13	16	7-10
K	149	166,32	12	1	13	2-3
L	136	116,85	13	5	18	13-14
M	376	88,86	6	10	16	7-10
N	320	94,95	8	9	17	11-12

Se observă în coloana 6 din tabelul 9.7 că ordonarea nu este satisfăcătoare întrucât există mai multe unități teritoriale cu acelaşi rang final.

Eliminarea acestei situații se poate face folosind ***metoda distanței relative față de performanța maximă***. În acest scop, unității teritoriale cu valoarea cea mai mare în coloana 2, tabelul 9.6, primeşte cifra 1,0000. Apoi, în ordine descrescătoare, celelalte unități teritoriale, primesc pe coloană valoarea procentuală din valoarea maximă 1489 u.m. care a primit cifra

Material cu drept de autor

1,0000. Astfel, prin aplicarea regulii de trei simplă: dacă pentru valoarea 1489 s-a primit cifra 1,0000, cât va primi procentual din 1489 mărimea 865 u.m. ? Ea va primi mărimea **x = (865/1489) = 0,5809**. În mod similar se procedează pentru toate valorile din coloana 2, tabelul 9.6, completându-se coloana 1 din tabelul 9.8. se ia în considerare apoi coloana 3 din tabelul 9.6 şi se completează la fel coloana 2 din tabelul 9.8.

Tabelul 9.8

A doua prelucrare a datelor din tabelul 9.6

Unita-tea	Distanța relativă față de performanță		Distanța medie	Rang final	Poziția (%) față de unitatea cea mai performantă
	Ptr. coloana 2 din tab. 9.6	Ptr. coloana 3 din tab. 9.6			
0	1	2	3	4	5
A	0,1988	0,5790	0,3393	10	43,67
B	0,5809	0,5176	0,5483	2	70,57
C	0,3056	0,4619	0,3757	5	48,35
D	0,2630	0,5305	0,3735	6	48,07
E	0,2283	0,6501	0,3853	4	49,59
F	0,0390	0,7999	0,1766	14	22,73
G	1,0000	0,6037	0,7770	1	100,00
H	0,1202	0,7405	0,2983	12	38,39
I	0,1632	0,7210	0,3430	9	44,14
J	0,4768	0,5028	0,4896	3	63,01
K	0,1001	1,0000	0,3164	11	40,72
L	0,0913	0,7026	0,2533	13	32,60
M	0,2525	0,5343	0,3673	7	47,27
N	0,2149	0,5709	0,3503	8	45,08

Distanța medie din coloana 3 din tabelul 9.8 reprezintă media geometrică a valorilor din coloanele 1 şi 2.

Se observă în tabelul 9.8 că distanța medie a fiecărei unități productive față de performanța maximă este cert diferită de la o unitate teritorială la alta, ceea ce permite, de data aceasta, o ierarhizare cât se poate de netă între ele.

Coloana 4 din tabelul 9.8 cuprinde noile ranguri finale pentru fiecare unitate teritorială, de la 1 la 14, câte un rang distinct pentru fiecare unitate teritorială, locul 1 revenind acelei unități care se situează cel mai aproape de performanța maximă, adică prezintă o „distanță medie" cât mai apropiată de 1, iar locul 14 se alocă acelei unități teritoriale care se află la distanța cea mai mare de performanța maximă. Coloana 5 din tabelul 9.8 conservă însă deosebirile dintre performanțele unităților teritoriale întrucât prezintă poziția

Material cu drept de autor

fiecărei unități în raport cu performanța maximă. Prin urmare, ultima coloană redă sub forma unei noi serii de mărimi relative de coordonare performanța medie a fiecărei unități teritoriale față de un maxim real găsit în colectivitatea cercetată.

În ierarhizarea multicriterială pot fi utilizate şi alte metode, printre care construcția unui indice complex de sinteză, cum a fost inicele dezvoltării umane prezentat anterior.

Exemplu de calcul 9.4. Pentru o societate comercială (S.C.) se cunosc datele statistice din tabelul 9.9 privind producția fizică (buc.), prețul unitar (u.m.) şi costuri unitare (u.m.).

Tabelul 9.9. Date statistice

S.C.	Producția fizică (buc)		Prețul unitar (u.m.)		Costuri unitare (u.m.)	
	q_0	q_1	p_0	p_1	c_0	c_1
A	1000	1250	100	150	80	90
B	500	800	250	300	45	50
C	90	270	175	200	50	60

Se cere:

a. Producția marfă, costurile şi profitul.
b. Rata rentabilității sau a profitului.
c. Indicii de grup şi abaterea absolută pentru costuri.
d. Indicele Paasche şi Laspeyres.
e. Indicele Marshall – Edgeworth.
f. Indicele Fisher.

Rezolvare.

a. Se realizează tabelul 9.10 pentru calcule intermediare:

Tabelul 9.10. Calcule intermediare pentru sume relevante

Firma	Producția marfă (u.m.)		Costuri totale (u.m.)		q_1c_0
	q_0p_0	q_1p_1	q_0c_0	q_1c_1	
0	1	2	3	4	5
A	100000	187500	80000	112500	100000
B	125000	240000	22500	40000	36000
C	15750	54000	4500	16200	13500
Total	240750	481500	107000	168700	149500

Tabelul 9.10 (continuare)

p_0q_1	p_1q_0	$p_1(q_1+q_0)$	$p_0(q_1+q_0)$	$q_1(p_0+p_1)$	$q_0(p_0+p_1)$
6	7	8	9	10	11
125000	150000	337500	225000	312500	250000

Material cu drept de autor

200000	150000	390000	325000	440000	275000
47250	18000	72000	63000	101250	33750
372250	318000	799500	613000	853750	558750

Producția marfă: $PM = \Sigma qp$; $PM_0 = \Sigma q_0 p_0 = 240750$;

$PM_1 = \Sigma q_1 p_1 = 481500$

Costurile totale: $CT = \Sigma qc$; $CT_0 = \Sigma q_0 c_0 = 107000$;

$CT_1 = \Sigma q_1 c_1 = 168700$

Profitul: $P = PM - CT$

$P_0 = PM_0 - CT_0 = \Sigma q_0 p_0 - \Sigma q_0 c_0 = 240750 - 107000 = 133750$

$P_1 = PM_1 - CT_1 = \Sigma q_1 p_1 - \Sigma q_1 c_1 = 481500 - 168700 = 312800$

b. Rata rentabilității (profitului).

$$Rp_0 = \frac{P_0}{C_0} = \frac{\Sigma q_0 p_0 - \Sigma q_0 c_0}{\Sigma q_0 c_0} = \frac{133750}{107000} = 1{,}25$$

$$Rp_1 = \frac{P_1}{C_1} = \frac{\Sigma q_1 p_1 - \Sigma q_1 c_1}{\Sigma q_1 c_1} = \frac{312800}{168700} = 1{,}85$$

c. Indicii de grup.

$$I_{1/0}^{qc} = \frac{\Sigma q_1 c_1}{\Sigma q_0 c_0} = \frac{168700}{107000} = 1{,}5766 \quad I_{1/0}^{q} = \frac{\Sigma q_1 c_0}{\Sigma q_0 c_0} = \frac{149500}{107000} = 1{,}397$$

$$I_{1/0}^{c} = \frac{\Sigma q_1 c_1}{\Sigma q_1 c_0} = \frac{168700}{149500} = 1{,}128$$

Proba: $I_{1/0}^{qc} = I_{1/0}^{q} \times I_{1/0}^{c} \Rightarrow 1{,}5766 \cong 1{,}397 \times 1{,}128 = 1{,}5758$

$\Delta_{1/0}^{qc} = \Sigma q_1 c_1 - \Sigma q_0 c_0 = 168700 - 107000 = 61700$

$\Delta_{1/0}^{q} = \Sigma q_1 c_0 - \Sigma q_0 c_0 = 149500 - 107000 = 42500$

$\Delta_{1/0}^{c} = \Sigma q_1 c_1 - \Sigma q_1 c_0 = 168700 - 149500 = 19200$

Proba: $\Delta_{1/0}^{qc} = \Delta_{1/0}^{q} + \Delta_{1/0}^{c} \Rightarrow 61700 = 42500 + 19200$

d. Indicele Paasche este indicele agregat în care ponderile sunt cele din perioada curentă:

$$P(p) = I_{1/0}^{p} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} = \frac{481500}{372250} = 1{,}29 \quad P(q) = I_{1/0}^{q} = \frac{\Sigma p_1 q_1}{\Sigma p_1 q_0} = \frac{481500}{318000} = 1{,}51$$

Material cu drept de autor

Indicele Laspeyres este indicele agregat în care ponderile sunt cele din perioada de bază:

$$L(p)=I_{1/0}^{p}=\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}=\frac{318000}{240750}=1{,}32 \qquad L(q)=I_{1/0}^{q}=\frac{\Sigma p_0 q_1}{\Sigma p_0 q_0}=\frac{372250}{240750}=1{,}546$$

e. Indicele Marshall – Edgeworth este un indice în care ponderile sunt din perioada de bază şi perioada curentă:

$$E(p)=\frac{\Sigma p_1(q_0+q_1)}{\Sigma p_0(q_0+q_1)}=\frac{799500}{613000}=1{,}304 \quad E(q)=\frac{\Sigma q_1(p_0+p_1)}{\Sigma q_0(p_0+p_1)}=\frac{853750}{558750}=1{,}5279$$

f. Indicele Fisher este indicele obținut ca medie geometrică a indicilor Paasche şi Laspeyres

$$F(p)=\sqrt{L(p)\times P(p)}=\sqrt{1{,}32\times 1{,}29}=1{,}3049$$

$$F(q)=\sqrt{L(q)\times P(q)}=\sqrt{1{,}546\times 1{,}51}=1{,}52789$$

Exemplu de calcul 9.5. O societate comercială (SC) realizează în perioada de bază (PB) şi perioada curentă (PC) datele statistice din tabelul 9.11.

Tabelul 9.11. Calcule intermediare

Produsul	Vânzări, mii €		Producție, mii t.		**Curs revenire, lei/€**	
	PB	**PC**	**PB**	**PC**	**PB**	**PC**
A	105	132	5	6	4600	8450
B	199,5	615	21	41	4700	8600

Se cere:

1. Indici, abateri, probe prin metoda restului nedescompus;
2. Indicii Paasche şi Laspeyres;
3. Indicele Marshall – Edgeworth.

Rezolvare.

Se realizează tabelul 9.12 pentru calcule intermediare pentru sume relevante.

Tabelul 9.12. Calcule intermediare pentru sume relevante

Produsul	Preț		Producție		Curs valutar	
	p_0	p_1	q_0	q_1	c_{r0}	c_{r1}
0	1	2	3	4	5	6
A	21	22	5	6	4600	8450
B	9,5	15	21	41	4700	8600
Total	-	-	-	-	-	-

Tabelul 9.12 (continuare)

$q_0p_0c_{r0}$	$q_1p_1c_{r1}$	$q_1p_0c_{r0}$	$q_0p_1c_{r0}$	$q_0p_0c_{r1}$
7	8	9	10	11
483000	1115400	579600	506000	887250
937650	5289000	1830650	1480500	1715700
1420650	6404400	2410250	1986500	2602950

Material cu drept de autor

1. Vânzări (€) = Producție (mii tone) x Preț → Preț (€/t) = = (Vânzări mii €)/Producție(mii tone).

Valoarea producției **VP** = Σ**pqc**$_r$.

$$I_{1/0}^{VP}=\frac{\Sigma q_1 p_1 c_{r1}}{\Sigma q_0 p_0 c_{r0}}=\frac{6404400}{1420650}=4{,}508 \quad I_{1/0}^{q}=\frac{\Sigma q_1 p_0 c_{r0}}{\Sigma q_0 p_0 c_{r0}}=\frac{2410250}{1420650}=1{,}697$$

$$I_{1/0}^{p}=\frac{\Sigma q_0 p_1 c_{r0}}{\Sigma q_0 p_0 c_{r0}}=\frac{1986500}{1420650}=1{,}398 \quad I_{1/0}^{c_r}=\frac{\Sigma q_0 p_0 c_{r1}}{\Sigma q_0 p_0 c_{r0}}=\frac{2602950}{1420650}=1{,}932$$

$$I_{1/0}^{q\cap p\cap c_r}=\frac{\Sigma q_1 p_1 c_{r1}}{\Sigma q_1 p_0 c_{r0}}\Big/\frac{\Sigma q_0 p_1 c_{r0}}{\Sigma q_0 p_0 c_{r0}}\Big/\frac{\Sigma q_0 p_0 c_{r1}}{\Sigma q_0 p_0 c_{r0}}=1{,}037$$

Proba:

$$I_{1/0}^{VP}=I_{1/0}^{q}\times I_{1/0}^{p}\times I_{1/0}^{c_r}\times I_{1/0}^{q\cap p\cap c_r}=4{,}508=1{,}697.1{,}398.1{,}832.1{,}037\cong 4{,}508$$

$$\Delta_{1/0}^{VP}=\Sigma q_1 p_1 c_{r1}-\Sigma q_0 p_0 c_{r0}=6404400-1420650=4983750$$

$$\Delta_{1/0}^{q}=\Sigma q_1 p_0 c_{r0}-\Sigma q_0 p_0 c_{r0}=2410250-1420650=989600$$

$$\Delta_{1/0}^{p}=\Sigma q_0 p_1 c_{r0}-\Sigma q_0 p_0 c_{r0}=1986500-1420650=565850$$

$$\Delta_{1/0}^{c_r}=\Sigma q_0 p_0 c_{r1}-\Sigma q_0 p_0 c_{r0}=2600950-1420650=1182300$$

$$\Delta_{1/0}^{q\cap d\cap p\cap c_r}=(\Sigma q_1 p_1 c_{r1}-\Sigma q_1 p_0 c_{r0})-(\Sigma q_0 p_1 c_{r0}-\Sigma q_1 p_0 c_{r0})-$$

$$-(\Sigma q_0 p_0 c_{r1}-\Sigma q_0 p_0 c_{r0})=3994150-565850-1182300=2246000$$

Proba:

$$\Delta_{1/0}^{VP}=\Delta_{1/0}^{q}+\Delta_{1/0}^{p}+\Delta_{1/0}^{c_r}+\Delta_{1/0}^{q\cap p\cap c_r}=4983750=989600+565850+$$

$$+1182300+2246000=4983750$$

Coeficienții de importanță sunt:

$$K_q=\frac{\Delta_{1/0}^{q}}{\Delta_{1/0}^{q}+\Delta_{1/0}^{p}+\Delta_{1/0}^{c_r}}=\frac{989600}{2737750}=0{,}361$$

$$K_p=\frac{\Delta_{1/0}^{p}}{\Delta_{1/0}^{q}+\Delta_{1/0}^{d}+\Delta_{1/0}^{p}+\Delta_{1/0}^{c_r}}=\frac{565850}{2737750}=0{,}207$$

$$K_{c_r}=\frac{\Delta_{1/0}^{c_r}}{\Delta_{1/0}^{q}+\Delta_{1/0}^{d}+\Delta_{1/0}^{p}+\Delta_{1/0}^{c_r}}=\frac{1182300}{2737750}=0{,}432$$

Abaterile absolute recalculate vor fi:

$$(\Delta_{1/0}^{q})^*=\Delta_{1/0}^{q}+K_q\times\Delta_{1/0}^{q\cap p\cap c_r}=989600+0{,}361\times 2246000=1800406$$

$$(\Delta_{1/0}^{p})^*=\Delta_{1/0}^{q}+K_p\times\Delta_{1/0}^{q\cap p\cap c_r}=565850+0{,}207\times 2246000=1030772$$

$$(\Delta_{1/0}^{c_r})^*=\Delta_{1/0}^{q}+K_{c_r}\times\Delta_{1/0}^{q\cap p\cap c_r}=1182300+0{,}432\times 2246000=2152572$$

Proba:

Material cu drept de autor

$$\Delta_{1/0}^{VP} = \Delta_{1/0}^{q} + \Delta_{1/0}^{p} + \Delta_{1/0}^{c_r} = 4983750 = 1800406 + 1030772 + 2152572 = 4983750$$

2. Indicii Paasche şi Laspeyres.

$$P(p) = \frac{\Sigma q_1 p_1}{\Sigma q_1 p_0} = \frac{747}{515,5} = 1,449; \quad P(q) = \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1} = \frac{747}{425} = 1,758$$

$$L(p) = \frac{\Sigma q_0 p_1}{\Sigma q_0 p_0} = \frac{425}{304,5} = 1,3957; \quad L(q) = \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} = \frac{515,5}{304,5} = 1,693$$

Tabelul 9.13. Sume relevante pentru calculul coeficienților **Paasche şi Laspeyres**.

Produs	Preț		Producție		q_0p_0	q_1p_1	q_1p_0	q_0p_1
	p_0	p_1	q_0	q_1				
A	21	22	5	6	105	132	110	126
B	9,5	15	21	41	1995	615	315	389,5
Total	-	-	26	47	3045	747	425	515,5

3. Indicele Marshall – Edgeworth.

Tabelul 9.14. Sume relevante pentru calculul indicelui **Marshall – Edgeworth**.

Produs	Preț		Producție		$p_0(q_0+q_1)$	$p_1(q_0+q_1)$	$q_0(p_0+p_1)$	$q_1(p_0+p_1)$
	p_0	p_1	q_0	q_1				
A	21	22	5	6	231	242	215	258
B	9,5	15	21	41	589	930	514,5	1004,5
Total	-	-	26	47	820	1172	7295	1262,5

$$E(p) = \frac{\Sigma p_1(q_0 + q_1)}{\Sigma p_0(q_0 + q_1)} = \frac{1172}{820} = 1,429; \quad E(q) = \frac{\Sigma q_1(p_0 + p_1)}{\Sigma q_0(p_0 + p_1)} = \frac{1262,5}{729,5} = 1,7306$$

Material cu drept de autor

CAP.10. ANALIZA STATISTICĂ A UNOR FENOMENE ECONOMICE SPECIFICE

10.1. Analiza statistică a productivității muncii

Productivitatea muncii exprimă sub formă sintetică eficiența cu care a fost cheltuită munca în procesul de producție.

Ca indicator de eficiență, productivitatea muncii se exprimă în forma sa cantitativă prin raportul dintre rezultatele (efectele) obținute într-un proces productiv şi efortul (cheltuiala de muncă) efectuat în respectivul proces sau prin inversul acestui raport.

Nivelul productivității muncii se exprimă, în mod sintetic, fie prin cantitatea de produse obținute într-o unitate de timp de muncă (**w**), fie prin cheltuiala de muncă ce revine pe unitatea de produs (**t**).

Corespunzător celor două posibilități de calcul a nivelului productivității muncii, în practică se utilizează:

- metoda directă:

$$w = \frac{q}{T} \tag{10.1}$$

- metoda inversă:

$$t = \frac{T}{q} \tag{10.2}$$

în care **q** este volumul producției; **T** – consumul de muncă în unități de timp.

Între cei doi indicatori ai productivității muncii există o relație de inversă proporționalitate:

$$w = \frac{1}{t}, \text{ respectiv: } t = \frac{1}{w}$$

Pentru măsurarea cât mai corectă a productivității muncii se pune problema, pe de o parte, a alegerii celei mai potrivite metode de calcul în funcție de nivelul organizatoric la care se face analiza, nivel la ce să se poată asigura cuantificarea cât mai exactă a celor doi indicatori: producție şi cheltuiala de timp de muncă ce trebuie să se refere la aceeaşi secțiune a procesului de producție, iar pe de altă parte, a alegerii celor mai potriviți indicatori pentru exprimarea volumului producției şi a cheltuielilor de timp de muncă.

Alegerea unei metode sau alteia trebuie să se facă în funcție de cerințele analizei statistice, de puterea de comparare a indicatorului la diferite nivele organizatorice sau de utilizarea lui în comparațiile internaționale.

De cele mai multe ori, în analiza statistică nu se foloseşte o singură metodă de calcul a nivelului productivității muncii, ci mai multe, pentru a crea posibilitatea alegerii acelor indicatori al căror nivel se apropie cel mai mult de nivelul real al acestuia.

Material cu drept de autor

Calculul nivelului şi dinamicii productivității muncii în funcție de modul de exprimare a producției. În scopul determinării nivelului productivității muncii, indicatorii producției industriale pot fi exprimați în unități naturale, natural convenționale, unități de timp de muncă şi valorice, iar timpul de muncă cheltuit poate fi exprimat prin număr de personal, număr mediu de muncitori, număr mediu de muncitori direct productivi, în om-zile şi om-ore. Ca indicator derivat, productivitatea muncii prezintă avantajele şi dezavantajele indicatorilor pe baza cărora se calculează. Semnificativi sunt indicatorii de rezultate (producția).

Indicatorul ***productivitatea muncii exprimat în unități naturale***, din punct de vedere teoretic, reflectă cel mai corect esența productivității muncii, deoarece caracterizează eficiența muncii direct din cantitatea de valori întrebuințare de acelaşi fel realizate într-o unitate de timp, sau ce volum de timp de muncă s-a cheltuit pentru obținerea unei unități de produs.

Folosirea acestui indicator este recomandată în toate cazurile posibile, în general în întreprinderile şi ramurile de producție omogenă, iar în ramurile cu producție eterogenă, pe feluri de produse.

Productivitatea muncii exprimată în unități naturale poate fi utilizată în analiza în dinamică sau în statică a nivelului productivității muncii pentru întreprinderi ce realizează aceleaşi produse sau pentru comparații pe plan internațional.

Deşi prezintă o serie de avantaje, posibilitățile de folosire a acestei metode sunt limitate.

Analiza variației în timp a productivității muncii se realizează diferit în funcție de metoda de calcul utilizată.

În cazul în care se utilizează metoda directă, nivelul productivității muncii ($\mathbf{w_i}$) reflectă volumul producției exprimat în unități fizice realizat în unitatea de timp.

Indicatorii dinamicii productivității muncii se vor calcula:

a. la nivelul unităților de producție omogenă sau pe tipuri de produse în cazul unei producții eterogene (dacă se pot evidenția cheltuielile de timp de muncă aferente):

$$i_{wi} = \frac{w_{i1}}{w_{i0}} \text{ sau } i_{wi} = R_{wi} + 1 \qquad (10.3)$$

$$R_{wi} = i_{wi} - 1 \text{ sau } R_{wi} = \frac{\Delta_{wi}}{w_{i0}} \qquad (10.4)$$

$$\Delta_{wi} = w_{i1} - w_{i0} \text{ sau } \Delta_{wi} = R_{wi}.w_{i0} \qquad (10.5)$$

$$i_{wi} > 1 \text{ dacă } i_{qi} > i_{Ti}$$

Creşterea productivității muncii în perioada curentă față de perioada de bază influențează pozitiv variația volumului producției:

Material cu drept de autor

$$\Delta_{q_i}^{w_i} = \Delta_{w_i} \times T_{i1}$$

b. la nivelul superior de agregare (grup de întreprinderi, ramură etc.) se calculează nivelul mediu al productivității muncii folosind una din relațiile:

$$\overline{w} = \frac{\sum_{i=1}^{n} q_i}{\sum_{i=1}^{n} T_i} = \frac{\sum_{i=1}^{n} w_i T_i}{\sum_{i=1}^{n} T_i} = \sum_{i=1}^{n} w_i y^{T_i} = \sum_{i=1}^{n} x_i^{\overline{w}} \qquad (10.6)$$

în care: **i = 1, 2,...,n** – varietatea întreprinderilor unde se realizează același produs; y^{T_i} - structura pe întreprinderi a timpului de muncă cheltuit; $x_i^{\overline{w}}$ - mărime cu care contribuie întreprinderea **i** la formarea nivelului mediu al productivității muncii.

Dinamica productivității medii este influențată de două grupe de factori:

- factori interni ce condiționează nivelul productivității muncii din fiecare întreprindere (progresul tehnic, factori legați de potențialul tehnic, de forța de muncă, de organizarea proceselor de producție etc.);
- factori externi de întreprindere ce evidențiază modificări în structura timpului de muncă.

Dinamica productivității medii se va analiza cu ajutorul indicelui cu structură variabilă ($I_{\overline{w}}$), indice ce reflectă influența simultană a celor două grupe de factori şi poate fi calculat:

$$I_{\overline{w}} = \frac{\overline{w}_1}{\overline{w}_0} = \frac{\sum q_{i1}}{\sum T_{i1}} : \frac{\sum q_{i0}}{\sum T_{i0}} = I_{\sum q} : I_{\sum T} \qquad (10.7)$$

respectiv:

$$I_{\overline{w}} = \frac{\overline{w}_1}{\overline{w}_0} = \frac{\sum w_{i1} T_{i1}}{\sum T_{i1}} : \frac{\sum w_{i0} T_{i0}}{\sum T_{i0}} = \frac{\sum w_{i1} y_1^{T_i}}{\sum w_{i0} y_0^{T_i}} = \frac{\sum x_1^{\overline{w}}}{\sum x_0^{\overline{w}}} \qquad (10.8)$$

Contribuția fiecărui factor se măsoară cu ajutorul indicilor factoriali:

- influența modificării productivității muncii din întreprinderile componente ale grupului este pusă în evidență de indicele cu structură fixă:

$$I_{\overline{w}}^{w} = \frac{\sum w_{i1} y_1^{T_i}}{\sum w_{i0} y_1^{T_i}} = \frac{\sum x_1^{\overline{w}}}{\sum x_*^{\overline{w}}} = \frac{\overline{w}_1}{\overline{w}_*} \qquad (10.9)$$

Material cu drept de autor

- influența modificării structurii timpului de muncă pe întreprinderi:

$$I_{\overline{w}}^{y^{T_i}} = \frac{\sum w_{i0} y_1^{T_i}}{\sum w_{i0} y_0^{T_i}} = \frac{\overline{w}_*}{\overline{w}_0} \qquad (10.10)$$

Modificarea absolută a productivității medii se va calcula:

$$\Delta_{\overline{w}_1} = \overline{w}_1 - \overline{w}_0 = \sum w_{i1} y_1^{T_i} - \sum w_{i0} y_o^{T_i} \qquad (10.11)$$

$$\Delta_{\overline{w}}^{w} = \overline{w}_1 - \overline{w}_* = \sum w_{i1} y_1^{T_i} - \sum w_{i0} y_1^{T_i} = \Delta_{w_i} y_1^{T_i} \qquad (10.12)$$

$$\Delta_{\overline{w}}^{y^T} = \overline{w}_* - \overline{w}_0 = \sum w_{i0} y_1^{T_i} - \sum w_{i0} y_o^{T_i} \qquad (10.13)$$

Prin multiplicarea cu ($\sum T_i$) a modificărilor absolute calculate se obține plusul sau minusul de producție:

$$\Delta_{\sum q}^{\overline{w}} = \Delta_{\overline{w}} . \sum T_{i1} \begin{cases} \Delta_{\sum_{T_i} q}^{w_i} = \Delta_{\overline{w}}^{w_i} . \sum T_{i1} \\ \Delta_{\sum_{T_i} q}^{y^{T_i}} = \Delta_{\overline{w}}^{y^{T_i}} . \sum T_{i1} \end{cases} \qquad (10.14)$$

Pentru calculul indicatorilor prezentați, se consideră cunoscute următoarele date din activitatea a trei întreprinderi producătoare de ciment pentru o lună din doi ani consecutivi, tabelul 10.1

Tabelul 10.1

Întreprinderea	Volumul producției (mii tone)		Timp de muncă (mii om-ore)		Productivitatea muncii (kg/om-oră)		$w_{i0}T_{i1}$ (mii tone)	I_{w_i} (%)	R_{w_i} (%)	Δ_{w_i} (kg/om-oră)
	0	1	0	1	0	1				
0	1	2	3	4	5	6	7	8	9	10
1	86,4	98	120	140	720	700	100,8	97,22	-2,78	-20
2	92,5	114,96	100	120	925	958	111	103,57	3,57	33
3	94,86	96,8	96	88	688,125	1100	86,755	111,32	11,32	118,875
Total	273,76	309,76	316	348	866,329	890,115	298,555	102,75	2,75	23,786

$$\overline{w}_0 = \frac{\sum q_{i0}}{\sum T_{i0}} = \frac{273,76 \times 10^6}{316 \times 10^6} = 866,329 \ \ kg/om - oră$$

Material cu drept de autor

$$\overline{w}_1 = \frac{\sum q_{i1}}{\sum T_{i1}} = \frac{309,76 \times 10^6}{348 \times 10^6} = 890,115 \ \ kg/om-oră$$

$$\overline{w}_* = \frac{\sum w_{i0} T_{i1}}{\sum T_{i1}} = \frac{298,555 \times 10^6}{348 \times 10^6} = 857,917 \ \ kg/om-oră$$

$$I_{\overline{w}} = \frac{\overline{w}_1}{\overline{w}_0} = \frac{890,115}{866,329} = 1,0275 \quad sau \quad 102,75\%$$

$$R_{\overline{w}} = I_{\overline{w}} - 1 = 0,0275 \quad sau \quad 2,75\%$$

$$\Delta_{\overline{w}} = \overline{w}_1 - \overline{w}_0 = 890,115 - 866,329 = 23,786 \, kg/om-oră$$

$$\Delta^{\overline{w}}_{\sum q} = \Delta_{\overline{w}} . \sum T_{i1} = 23,786 \times 348 \times 10^3 = 8277,5 \, tone$$

$$I^{\overline{w}}_{\overline{w}} == \frac{\overline{w}_1}{\overline{w}_*} = \frac{890,115}{857,917} = 1,03753 \ sau \ 103,753\%$$

$$R^{w_1}_{\overline{w}} = I^{w_1}_{\overline{w}} - 1 = 1,03753 - 1 = 0,03753 \quad sau \ 3,753\%$$

$$\Delta^{w}_{\overline{w}} = \overline{w}_1 - \overline{w}_* = 890,115 - 857,917 = 32,198 \ kg/om-oră$$

$$\Delta^{\overline{w}_1}_{\sum q_1} = \Delta^{w_1}_{\overline{w}} . \sum T_{i1} = 32,198 \times 348 \times 10^3 = 11204,9 \, tone$$

$$I^{\overline{w}_*}_{\overline{w}} = \frac{857,917}{866,329} = 0,99029 \quad sau \quad 99,029\%$$

$$R^{y^{T_1}}_{\overline{w}} = I^{y^{T_1}}_{\overline{w}} - 1 = -0,00971 \ \ sau \ \ -0,971\%$$

$$\Delta^{y^{T_1}}_{\overline{w}} = \overline{w}_* - \overline{w}_0 = 857,917 - 866,329 = -8,412 \, kg/om-oră$$

$$\Delta^{T_1}_{\sum q_1} = \Delta^{y^{T_1}}_{\overline{w}} . \sum T_{i1} = -8,412 \times 348 \times 10^3 = -2927,3 \, tone$$

Din analiza rezultatelor se constată o sporire a nivelului productivității muncii la unitățile 2 şi 3 cu 3,57%, respectiv cu 11,39%, iar şa întreprinderea 1 s-a înregistrat o scădere a nivelului acesteia cu 2,78%, respectiv 20 kg/om-oră.

La nivelul celor trei întreprinderi, nivelul mediu al productivității muncii a crescut cu 2,75%, respectiv 23,786 kg/om-oră, fapt ce a contribuit la realizarea unui spor de producție de 8277,5 tone.

Material cu drept de autor

Pe factori, situația se prezintă diferit. Creşterea nivelului mediu al productivității muncii din cele trei întreprinderi a fost de 3,753%, respectiv 32,298 kg/om-oră şi de producție de 11204,9 tone.

Variația structurii timpului a determinat o reducere a nivelului productivității muncii cu 0,971%, respectiv 8,412 kg/om-oră şi de producție de 2927,3 tone.

Productivitatea muncii calculată sub forma timpului de muncă consumat pe unitate de produs (metoda inversă), este utilizată tot mai mult în analiza statistică întrucât permite: identificarea şi evaluarea căilor de creştere a productivității muncii; determinarea necesarului de personal; analiza corelată cu gradul îndeplinirii normelor cu randamentul utilajului, cu cheltuielile, cu salariile ce revin pe unitatea de produs etc.

Se recomandă utilizarea indicatorului ***timp consumat pe unitatea de produs*** pentru produsele complexe pe operații, respectiv faze ale procesului de producție. Pentru aceasta se impune stabilirea consumului total de timp de muncă aferent unui produs, luându-se în calcul manopera directă şi cea indirectă. Repartizarea manoperei indirecte pe produs se face pe baza unor chei de repartizare.

Analiza în dinamică a timpului consumat pe unitatea de produs se poate realiza:

a. la nivelul întreprinderii cu producție omogenă (respectiv pe tip de produs în întreprinderile cu producție eterogenă) şi în acest caz se calculează:

- indicele dinamicii timpului consumat pe unitatea de produs ($\mathbf{i_{ti}}$):

$$i_{ti} = \frac{t_{i0}}{t_{i1}} \qquad (10.15)$$

Calculat sub forma (10.15), el este egal cu indicele productivității muncii calculat după metoda directă:

$$i_{ti} = \frac{T_{i0}}{q_{i0}} : \frac{T_{i1}}{q_{i1}} = \frac{T_{i0}}{T_{i1}} : \frac{q_{i0}}{q_{i1}} = \frac{1}{i_{T_i}} : \frac{1}{q_i} = \frac{i_{q_i}}{i_{T_i}} = i_{wi} \qquad (10.16)$$

- modificarea relativă ($\mathbf{R_{ti}}$):

$$R_{ti} = i_{ti} - 1 \qquad (10.17)$$

sau:

$$R_{ti} = \frac{t_{i0}}{t_{i1}} - 1 = \frac{t_{i0} - t_{i1}}{t_{i1}} = \frac{\Delta_{ti}}{t_{i1}} \qquad (10.18)$$

- modificarea absolută (Δ_{ti}):

$$\Delta_{ti} = t_{i0} - t_{i1} \qquad (10.19)$$

respectiv:

Material cu drept de autor

$$\Delta_{ti} = R_{ti}.t_{i1} \qquad (10.20)$$

- plusul sau minusul de producție realizat ca urmare a modificării timpului pe unitate de produs:

$$\Delta_{qt}^{ti} = (t_{i0} - t_{i1}).q_{i1} : t_{i0} \qquad (10.21)$$

b. la nivel superior de agregare (grup de întreprinderi, ramură) se calculează timpul mediu pe unitate de produs (**i**):

$$\bar{t} = \frac{\sum T_i}{\sum q_i} \qquad (10.21)$$

respectiv:

$$\bar{t} = \frac{\sum t_i q_i}{\sum q_i} = \sum t_i y^{q_t} = \sum x_i^{\bar{t}} \qquad (10.22)$$

în care y^{q_t} este structura pe întreprinderi a producției; $x_i^{\bar{t}}$ - mărimea cu care contribuie fiecare întreprindere la formarea timpului mediu pe unitatea de produs.

Factorii ce influențează modificarea în timp a indicatorului sunt: $\mathbf{t_i}$ – timpul consumat pe unitatea de produs la nivelul întreprinderii **i** (factor intern de întreprindere); y^{q_t} - structura producției (factor extern de întreprindere).

Dinamica ***timpului mediu consumat pe unitatea de produs*** se va calcula cu ajutorul indicelui:

$$I_{\bar{t}} = \frac{\bar{t}_{ti}}{\bar{t}_i} \text{ sau } I_{\bar{t}} = \frac{\sum t_{i0} q_{i0}}{\sum q_{i0}} : \frac{\sum t_{i1} q_{i1}}{\sum q_{i1}} = \frac{\sum t_{i0} y_0^{q_t}}{\sum t_{i1} y_1^{q_t}} \qquad (10.23)$$

Contribuția factorilor este pusă în evidență de indicii factoriali:
- influența modificării lui $\mathbf{t_i}$:

$$I_{\bar{t}}^{t_i} = \frac{\sum t_{i0} y_1^{q_t}}{\sum t_{i1} y_1^{q_t}} = \frac{\bar{t}_*}{\bar{t}_1} \qquad (10.24)$$

- influența modificării lui y^{q_t} :

$$I_{\bar{t}}^{q_t} = \frac{\sum t_{i0} y_0^{q_t}}{\sum t_{i0} y_1^{q_t}} = \frac{\bar{t}_0}{\bar{t}_*} \qquad (10.25)$$

Între cei trei indici există relația:

$$I_{\bar{t}} = I_{\bar{t}}^{t_i} \times I_{\bar{t}}^{y^{q_t}} \qquad (10.26)$$

Material cu drept de autor

Modificarea absolută a timpului mediu consumat pe unitatea de produs se calculează, în general, şi pe factori, ca diferență între numărătorul şi numitorul indicilor corespunzători.

Plusul sau minusul de producție datorat variației ***timpului mediu consumat pe unitatea de produs*** se va calcula astfel:

$$\Delta^{\bar{t}}_{\sum q_t} = \left(\bar{t}_0 - \bar{t}_1\right)\sum q_{i1} : \bar{t}_0 \qquad (10.27)$$

$$\Delta^{t_i}_{\sum q_t} = \left(\bar{t}_* - \bar{t}_1\right)\sum q_{i1} : \bar{t}_0 \qquad (10.28)$$

$$\Delta^{y^{q_t}}_{\sum q_t} = \left(\bar{t}_0 - \bar{t}_*\right)\sum q_{i1} : \bar{t}_0 \qquad (10.29)$$

Între modificările calculate se stabileşte relația:

$$\Delta^{\bar{t}}_{\sum q_t} = \Delta^{t_i}_{\sum q_t} + \Delta^{y^{q_t}}_{\sum q_t} \qquad (10.30)$$

Pentru un agent economic care dispune de mai multe unități care realizează acelaşi produs (unități producătoare de energie electrică, de exploatare a cărbunelui, a țițeiului etc.) prin compararea indicatorilor calculați cu relațiile de mai sus se pot desprinde concluzii cu privire la factorii ce trebuie activați în sensul reducerii timpului consumat pe unitatea de produs ($\mathbf{t_i}$), respectiv a redistribuirii producției în favoarea unităților ce realizează un timp ce revine pe unitatea de produs mai mic.

Pentru calculul şi analiza indicatorilor prezentați mai sus se vor utiliza datele din tabelul 10.1 cu privire la volumul producției şi a timpului de muncă cheltuit (coloanele 1, 2, 3, 4), realizându-se tabelul 10.2.

Tabelul 10.2

Întreprinderea	Timp ce revine pe tona de ciment, (ore/tonă)		$t_{i0}q_{it}$ (mii om-ore)
	0	1	
0	1	2	3
1	1,389	1,429	136,122
2	1,081	1,044	124,272
3	1,012	0,909	97,962
Total	1,1543	1,1235	358,356

$$\bar{t}_0 = \frac{\sum T_{i0}}{\sum q_{i0}} = \frac{316 \times 10^3}{273{,}76 \times 10^3} = 1{,}1543\, ore/tonă$$

$$\bar{t}_1 = \frac{\sum T_{i1}}{\sum q_{i1}} = \frac{348 \times 10^3}{309{,}76 \times 10^3} = 1{,}1235\, ore/tonă$$

Material cu drept de autor

$$\bar{t}* = \frac{\sum t_{i0} q_{i1}}{\sum q_{i1}} = \frac{358{,}356 \times 10^3}{309{,}76 \times 10^3} = 1{,}1688\ ore / tonă$$

$$I_{\bar{t}} = \frac{\bar{t}_0}{\bar{t}_1} = \frac{1{,}1543}{1{,}1235} = 1{,}0274 \ \ sau\ 102{,}74\%$$

$$\Delta^{\bar{t}}_{\sum q_t} = \left(\bar{t}_0 - \bar{t}_1\right) \sum q_{i1} : \bar{t}_0 =$$

$$= (1{,}1543 - 1{,}1235) \times 309{,}76 \times 10^3 : 1{,}1543 = 8265{,}3\ tone$$

$$I^{t_i}_{\bar{t}} = \frac{\sum t_{i0} y_1^{q_t}}{\sum t_{i1} y_1^{q_t}} = \frac{\bar{t}*}{\bar{t}_1} = \frac{1{,}15688}{1{,}1235} = 1{,}0297 \ \ sau\ 102{,}97\%$$

$$\Delta^{t_i}_{\sum q_t} = \left(\bar{t}* - \bar{t}_1\right) \sum q_{i1} : \bar{t}_0 =$$

$$= (1{,}15688 - 1{,}1235) \times 309{,}76 \times 10^3 : 1{,}1543 = 8957{,}6\ tone$$

$$I^{q_t}_{\bar{t}} = \frac{\sum t_{i0} y_0^{q_t}}{\sum t_{i0} y_1^{q_t}} = \frac{\bar{t}_0}{\bar{t}*} = \frac{1{,}1543}{1{,}15688} = 0{,}99776\ sau\ 99{,}776\%$$

$$\Delta^{y^{q_t}}_{\sum q_t} = \left(\bar{t}_0 - \bar{t}*\right) \sum q_{i1} : \bar{t}_0 =$$

$$= (1{,}1543 - 1{,}15688) \times 309{,}76 \times 10^3 : 1{,}1543 = -0692{,}3\ tone$$

La nivelul celor trei întreprinderi s-a înregistrat o reducere a timpului consumat pe tona de ciment, fapt ce a contribuit la realizarea unui spor de producție de 8265,3 tone. Această modificare s-a datorat reducerii timpului consumat pe tona de ciment la întreprinderile 2 şi 3 (8957,6 tone) şi influenței nefavorabile a structurii producției (-0692,3 tone).

Calculul nivelului şi dinamicii productivității muncii exprimată în unități natural convenționale. Indicatorul producției exprimat în unități natural convenționale serveşte la calculul productivității muncii exprimată în unități natural convenționale şi are o sferă de aplicabilitate mai largă decât în cazul exprimării productivității muncii în unități naturale.

Esența acestei metode constă în transformarea unor produse calitativ deosebite într-un produs convențional cu ajutorul unor coeficienți de transformare (**K**).

În analiza statistică se recomandă utilizarea acestui indicator în toate cazurile posibile şi folosirea lui în comparațiile efectuate în timp şi în spațiu.

Productivitatea muncii se calculează în felul următor:

a. La nivelul unității:

Material cu drept de autor

$$w_j = \frac{\sum_{i=1}^{n} q_{ij} K_i}{T_j} \qquad (10.31)$$

b. la nivel superior de agregare:

$$\overline{w} = \frac{\sum_{j=1}^{m}\sum_{i=1}^{n} q_{ij} K_i}{\sum T_j} \qquad (10.32)$$

sau:

$$\overline{w} = \frac{\sum_{j} w_j T_j}{\sum T_j} = \sum_{j} w_j y^{T_j} \qquad (10.33)$$

în care **i = 1, 2,...n** este varietatea tipurilor calitative; **j = 1, 2,...,m** – varietatea întreprinderilor economice ce realizează aceleaşi tipuri calitative; $\mathbf{w_j}$ – productivitatea medie a muncii exprimată în unităţi natural convenţionale din întreprinderea **j**; $\mathbf{K_i}$ – coeficientul de transformare corespunzător tipului calitativ **i**; $\sum_{i=1}^{n} q_{ij} K_i$ - producţia exprimată în unităţi natural convenţionale la întreprinderea **j**: $\mathbf{T_j}$ – timp de muncă cheltuit (poate fi exprimat prin număr mediu de personal, număr mediu de muncitori, om-ore sau om-zile); $\overline{w}$ - productivitatea medie a muncii exprimată în unităţi natural convenţionale calculată la nivelul unui grup de întreprinderi; $\sum_{j=1}^{m}\sum_{i=1}^{n} q_{ij} K_i$ - volumul producţiei exprimat în unităţi natural convenţionale determinat la nivelul gru8pului de întreprinderi; $\sum_{j=1}^{m} T_j$ - volumul de timp de muncă cheltuit, determinat la nivelul grupului de întreprinderi.

Dinamica productivităţii muncii în acest caz se va calcula:

a. la nivelul unităţilor:

$$i_{w_j} = \frac{w_{j1}}{w_{j0}} \qquad (10.34)$$

Material cu drept de autor

b. la nivel superior de agregare:

$$I_{\overline{w}} = \frac{\overline{w}_1}{\overline{w}_0} = \frac{\sum_{j=1}^{m} w_{j1} y_1^{T_j}}{\sum_{j=1}^{m} w_{j0} y_0^{T_j}} \qquad (10.35)$$

Pentru analiza influențelor factorilor în cifre relative şi absolute se va utiliza sistemul de indicatori prezentat la dinamica productivității muncii exprimată în unități naturale, calculată după metoda directă.

Calculul nivelului şi dinamicii productivității muncii exprimată în unități valorice. Metoda valorică de exprimare a producției industriale prezintă o serie de ***avantaje*** față de celelalte metode de evidență a producției industriale şi poate fi aplicată la toate nivelurile organizatorice.

În întreprinderile industriale analiza nivelului şi dinamicii productivității muncii se calculează pe baza diferiților indicatori valorici ai producției industriale (producția marfă, cifra de afaceri, valoarea adăugată, producția globală etc.).

Pentru calculul nivelului productivității muncii, indicatorii valorici se exprimă în prețuri curente (acestea sunt prețurile produselor industriale în primul stadiu al comercializării fără TVA).

Analiza în dinamică a productivității muncii impune recalcularea indicatorilor din prețuri curente în prețuri comparabile.

La nivelul întreprinderilor, productivitatea muncii se calculează astfel:

$$w = \frac{Q(Pm, Ca, Va, Pg, etc.)}{T} \qquad (10.36)$$

Productivitatea muncii se mai poate calcula şi cu ajutorul unor ***modele multiplicative*** de forma:

$$w_{Va} = \frac{Va}{T} = \frac{Pg}{T} \times \frac{Va}{Pg} \qquad (10.37)$$

$$w_{Pa} = \frac{Pm}{T} = \frac{Pg}{T} \times \frac{Pm}{Pg} \qquad (10.38)$$

în care w_{Va} este productivitatea muncii calculată pe baza valorii adăugate; **Va** – valoarea adăugată; **T** – timpul de muncă cheltuit; **Pg** – producția globală; w_{Pm} – productivitatea muncii calculată pe baza producției marfă fabricate; **Pm** – producția marfă fabricată.

Analizând dinamica indicatorilor calculați cu ajutorul relațiilor de mai sus se pot desprinde o serie de ***concluzii***, şi anume:

Material cu drept de autor

- ritmul de creştere al valorii adăugate ce revine pe o persoană este mai mare decât ritmul de creştere al producției globale ce revine pe o persoană dacă dinamica valorii adăugate devansează dinamica producției globale ($I_{Va} > I_{Pg}$). Aceasta se realizează în cazul în care consumul intermediar înregistrează o reducere de timp;
- dinamica producției marfă ce revine pe o persoană devansează dinamica producției globale ce revine pe o persoană dacă ritmul de creştere al producției marfă este superior ritmului de creştere al producției globale, situație ce se înregistrează în cazul în care stocurile de producție neterminată şi semifabricate au o tendință de reducere.

Pentru analiza variației productivității muncii calculate pe baza unor indicatori valorici ai producției se utilizează datele din tabelul 10.3.

Tabelul 10.3

Nr. crt	Indicatori	Sim-bol	Unități monetare	Perioada		$I_{1/0}$ (%)	$R_{1/0}$ (%)
				de bază	curentă		
1	Prod. marfă	Pm	mil.lei/ persoană	750	795	106	6
2	Cifra de afaceri	Ca	mil.lei/ persoană	675	719,55	106,6	6,6
3	Valoarea adăugată	Va	mil.lei/ persoană	367,021	394,839	107,5	7,5
4	Prod. globală	Pg	mil.lei/ persoană	797,872	822,581	102,5	2,5
5	Numărul mediu de personal	T	nr.	200	205	102,5	2,5
6	Productivitatea medie lunară pe baza:						
	- producției marfă;	w_{Pm}	mii lei/ persoană	3750	3878,048	103,41	3,41
	- cifrei de afaceri;	w_{Ca}	mii lei/ persoană	3375	3510	104	4
	- valorii adăugate;	w_{Va}	mii lei/ persoană	1835,105	1926,044	104,96	4,96
	- producției globale.	w_{Pg}	mii lei/ persoană	3989,36	4012,59	100,58	0,58

Indicatorii valorici se referă la o perioadă de două luni, cu acelaşi număr de zile lucrătoare, calculați în prețuri comparabile.

Din analiza indicatorilor în tabelul 10.3, rezultă că nivelul productivității muncii a înregistrat creşteri pentru toate variantele de calcul. Ritmurile de creştere însă, sunt diferite:

$$R_{w_{Pm}} > R_{w_{Pg}}$$

Material cu drept de autor

Având în vedere faptul că productivitatea muncii calculată pe baza producției marfă reflectă volumul producției destinat circuitului economic ce revine pe o persoană, dinamica acesteia trebuie să devanseze dinamica producției globale ce revine pe o persoană – indicator al cărui nivel este influențat de variația stocurilor de semifabricate şi de producție neterminată.

De asemenea, ritmul de creştere al productivității muncii calculat pe baza cifrei de afaceri devansează ritmul productivității uncii calculat pe baza producției marfă, situație ce reflectă o creştere a volumului vânzărilor produselor fabricate şi o diminuare a volumului producției rămase în stoc.

Analizând creşterea valorii adăugate ce revine pe o persoană şi a producției globale ce revine pe o persoană, se constată o situație pozitivă: ($R_{w_{Va}} > R_{w_{Pg}}$), fapt ce evidențiază o reducere în timp a ponderii consumului intermediar în producția globală.

Întrucât indicatorii valorici ai producției se calculează ca sumă a elementelor componente la nivelul întreprinderii (nu şi a verigilor organizatorice), productivitatea muncii are caracter de medie ($\mathbf{w_i}$), iar la nivelul grupului de întreprinderi, ramură etc., productivitatea medie ($\overline{w}$) se va calcula ca medie a acestor productivități:

$$\overline{w} = \frac{\sum_{i=1}^{n} Q_i}{\sum_{i=1}^{n} T_i} \qquad (10.39)$$

sau:

$$\overline{w} = \frac{\sum_{i=1}^{n} w_i T_i}{\sum_{i=1}^{n} T_i} = \sum_{i=1}^{n} w_i y_i^{T_i} = \sum_{i=1}^{n} x_i^{\overline{w}} \qquad (10.40)$$

în care **i = 1, 2,…,n** este varietatea întreprinderilor componente ale grupului (ramurii); $\mathbf{Q_i}$ – un indicator valoric al producției din întreprinderea **i**; $\mathbf{T_i}$ – timp de muncă cheltuit în întreprinderea **i**; $\mathbf{w_i}$ – productivitatea muncii la întreprinderea **i**; $\mathbf{y^{Ti}}$ – greutatea specifică a timpului de muncă cheltuit în întreprinderea **i** în timpul total cheltuit la nivelul grupului de întreprinderi.

Factorii ce influențează dinamica productivității medii sunt:

- productivitatea medie a muncii calculată la nivelul unităților;
- structura timpului de muncă cheltuit.

Dinamica productivității medii se calculează cu ajutorul indicilor de grup sub formă de raport a două medii:

Material cu drept de autor

$$I_{\overline{w}} = \frac{\overline{w}_1}{\overline{w}_0} = \frac{\sum_{i=1}^{n} Q_{i1}}{\sum_{i=1}^{n} T_{i1}} : \frac{\sum_{i=1}^{n} Q_{i0}}{\sum_{i=1}^{n} T_{i0}} = I_{\sum_{i=1}^{n} Q_i} : I_{\sum_{i=1}^{n} T_i} \qquad (10.41)$$

sau:

$$I_{\overline{w}} = \frac{\overline{w}_1}{\overline{w}_0} = \frac{\sum_{i=1}^{n} w_{i1} T_{i1}}{\sum_{i=1}^{n} T_{i1}} : \frac{\sum_{i=1}^{n} w_{i0} T_{i0}}{\sum_{i=1}^{n} T_{i0}} = \frac{\sum_{i=1}^{n} w_{i1} y_1^{T_1}}{\sum_{i=1}^{n} w_{i0} y_0^{T_0}} = \frac{\sum_{i=1}^{n} x_i^{\overline{w}_1}}{\sum_{i=1}^{n} x_i^{\overline{w}_0}} \qquad (10.42)$$

Indicii factoriali şi modificările absolute corespunzătoare se calculează la fel ca şi în cazul productivității muncii exprimată în unități naturale.

Pentru exemplificarea se vor utiliza datele prezentate în tabelul 10.4.

Tabelul 10.4

Întreprinderi	Producția marfă (mil. lei prețuri comparabile)		Numărul mediu de salariați	
	III/2000	IV/2000	III/2000	IV/2000
0	1	2	3	4
1	192	306	48	68
2	560	530,4	112	102
Total	752	836,4	160	170
Notă: Datele au fost calculate pentru acelaşi număr de zile lucrătoare.				

Pe baza datelor din tabelul 10.4 s-a calculat productivitatea medie a muncii la nivelul unităților şi pe total. Aceşti indicatori se regăsesc în tabelul 10.5.

Tabelul 10.5

Înt.	Productivitatea medie a muncii, mii lei pe salariat		I_{wi}, (%)	R_{wi}, (%)	Δ_{wi}, mii lei/ salariat	Structura numărului mediu de personal (%)		$x_i^{\overline{w}} = w_i y^{T_i}$ (mii lei)		
	III/2000	IV/2000				III/2000	IV/2000			
0	1	2	3	4	5	6	7	8	9	10
1	4500	4500	112,5	12,5	500	30	40	1200	1800	1600
2	5000	5200	104	4	200	70	60	3500	3120	3000
Total	4700	4920	104,681	4,681	220	100	100	4700	4920	4600

Material cu drept de autor

Productivitatea muncii a crescut cu 12,5% la unitatea 1 şi cu 4% la unitatea 2. Pe total, nivelul productivității muncii a crescut cu 4,681%, respectiv 220 mii lei/salariat.

$$I_{\overline{w}} = \frac{\overline{w}_1}{\overline{w}_0} = \frac{4920 \times 10^3}{4700 \times 10^3} = 1{,}04681$$

$$\Delta_{\overline{w}} = \overline{w}_1 - \overline{w}_0 = (4920 - 4700) \times 10^3 = 220\, mii\, lei / salariat$$

Modificarea productivității muncii din fiecare întreprindere a determinat o creştere cu 6,59%, respectiv 320 mii lei/salariat.

$$I_{\overline{w}}^{w_i} = \frac{\sum_{i=1}^{2} w_{i1} y_1^{T_i}}{\sum_{i=1}^{2} w_{i0} y_1^{T_i}} = \frac{4920 \times 10^3}{4600 \times 10^3} = 1{,}06956$$

$$\Delta_{\overline{w}}^{w_i} = \overline{w}_1 - \overline{w}_* = (4920 - 4600) \times 10^3 = 320\, mii\, lei / salariat$$

Variația structurii numărului mediu de salariați a influențat negativ dinamica productivității medii, contribuind la reducerea nivelului acesteia cu 2,128%, respectiv 100 de mii lei/salariat ca urmare a faptului că la unitatea 2 a scăzut în perioada curentă ponderea numărului mediu de salariați la 60% față de 70% cât a fost în perioada de bază. Aceasta s-a înregistrat tocmai la unitatea care a realizat un nivel mai ridicat al productivității muncii:

$$I_{\overline{w}}^{y^{T_i}} = \frac{\sum_{i=1}^{2} w_{i0} y_1^{T_i}}{\sum_{i=1}^{2} w_{i0} y_0^{T_i}} = \frac{4600 \times 10^3}{4700 \times 10^3} = 0{,}97872$$

$$\Delta_{\overline{w}}^{y^{T_i}} = \overline{w}_* - \overline{w}_0 = (4600 - 4700) \times 10^3 = -100\, mii\, lei / salariat$$

Rezultatele obținute privind influența factorilor verifică relațiile dintre:

- indicele general şi indicii factoriali:

$$I_{\overline{w}} = I_{\overline{w}}^{w_i} \times I_{\overline{w}}^{y^{T_i}} = 1{,}04681 = 1{,}06956 \times 0{,}97872$$

- modificarea absolută totală şi influențele absolute datorate variației celor doi factori:

$$\Delta_{\overline{w}} = \Delta_{\overline{w}}^{w_i} + \Delta_{\overline{w}}^{y^{T_i}} = 220 \times 10^3 = 320 \times 10^3 - 100 \times 10^3$$

Material cu drept de autor

Calculul nivelului şi dinamicii productivității muncii exprimată în unități de timp de muncă. În determinarea nivelului şi dinamicii productivității muncii, ***metoda unităților de timp de muncă prezintă un interes deosebit, ca urmare a faptului că permite cuantificarea productivității muncii nu numai la nivelul întreprinderii, ci şi al structurilor organizatorice, iar în cadrul acestora pe formații de lucru şi chiar al locurilor de muncă***.

Astfel, pot fi identificate rezervele de creştere ale productivității muncii şi se pot compara rezultatele obținute cu sarcinile de producție precum şi ci cheltuielile cu manopera.

Aplicarea metodei prezintă însă o serie de dificultăți în determinarea volumului de timp de muncă aferent producției realizate. Determinat pe baza timpului normat pe unitatea de produs, în calculul productivității muncii se ia doar timpul de muncă cheltuit de muncitorii ce lucrează în acord. Acest neajuns se poate înlătura în cazul în care se urmăreşte consumul de manoperă pe faze ale procesului tehnologic pentru muncitorii care lucrează în acord şi prin redistribuirea timpului de muncă cheltuit pentru celelalte categorii de personal pe baza unor chei de repartizare.

Productivitatea muncii este dată de relația:

$$w = \frac{\sum_{i=1}^{n} q_i t_{ni}}{\sum_{i=1}^{n} T_i} \qquad (10.43)$$

în care **w** este productivitatea muncii exprimată în unități de timp de muncă; $\mathbf{q_i}$ – volumul fizic al producției pe sortimente, faze, operații etc.; $\mathbf{t_{ni}}$ – timp normat pe unitatea de produs, fază, operație etc.; $\mathbf{T_i}$ – timp de muncă cheltuit pentru obținerea producției $\mathbf{q_i}$ exprimat în ore-om.

Timpul de muncă cheltuit ($\mathbf{T_i}$) se poate calcula:

$$T_i = q_i . t_i$$

în care $\mathbf{t_i}$ este timp efectiv ce revine pe unitatea de produs.

Dinamica productivității muncii se calculează astfel:

$$I_w = \frac{\sum_{i=1}^{n} q_{i1} . t_{ni}}{\sum_{i=1}^{n} T_{i1}} : \frac{\sum_{i=1}^{n} q_{i0} t_{ni}}{\sum_{i=1}^{n} T_{i0}} \qquad (10.44)$$

sau:

Material cu drept de autor

$$I_w = \frac{\sum_{i=1}^{n} q_{i1} . t_{ni}}{\sum_{i=1}^{n} q_{i1} t_{i1}} : \frac{\sum_{i=1}^{n} q_{i0} t_{ni}}{\sum_{i=1}^{n} q_{i0} t_{i0}} \qquad (10.45)$$

Analiza dinamicii productivității muncii în cazul în care producția este exprimată în unități de timp de muncă este condiționată de calitatea normelor de timp. În consecință, acestea trebuie să fie determinate în mod ştiințific.

În calculul dinamicii productivității muncii, ***timpul normat pe unitatea de produs*** se utilizează ca element de omogenizare a producției şi se ia constant în ambele perioade.

Factorii care influențează dinamica productivității muncii sunt:

- timpul consumat pe unitatea de produs ($\mathbf{t_i}$);
- volumul şi structura producției (y^{q_i}).

Influența separată a factorilor se calculează cu ajutorul indicilor:

$$I_w^{t_i} = \frac{\sum q_{i1} t_{ni}}{\sum q_{i1} t_{i1}} : \frac{\sum q_{i1} t_{ni}}{\sum q_{i1} t_{i0}} = \frac{\sum q_{i1} t_{i0}}{\sum q_{i1} t_{i1}} \qquad (10.46)$$

Din punct de vedere economic există o situație bună în cazul realizării inegalității: $I_w^{t_i} > 1$.

Diferența dintre numărătorul şi numitorul indicelui evidențiază timpul economisit ($\Delta > 0$) sau cheltuit în plus ($\Delta < 0$) pentru producția realizată în perioada de analiză ca urmare a modificării lui $\mathbf{t_i}$:

$$I_w^{y^{q_i}} = \frac{\sum q_{i1} t_{ni}}{\sum q_{i1} t_{i0}} : \frac{\sum q_{i0} t_{ni}}{\sum q_{i0} t_{i0}} \qquad (10.47)$$

Calculul nivelului şi dinamicii productivității muncii în funcție de modul de exprimare a cheltuielilor de timp de muncă. Determinarea cât mai corectă a indicatorilor ce reflectă cheltuiala de timp de muncă pentru producție trebuie să constituie o preocupare permanentă a factorilor de decizie din întreprinderi.

Cheltuiala de timp de muncă este, de fapt, indicatorul de efort utilizat în calculul nivelului productivității muncii.

Analiza modificărilor intervenite în consumul de muncă oferă posibilitatea descoperirii şi mobilizării rezervelor de creştere a productivității muncii.

Pentru cuantificarea nivelului productivității muncii se utilizează indicatori ai numărului de personal (total personal, muncitori, muncitori direct productivi) şi ai volumului de timp de muncă exprimați ore-om şi zile-om.

Material cu drept de autor

Ca număr de personal poate fi luat numărul mediu (scriptic) sau numărul mediu efectiv.

Din analiza conținutului indicatorului numărului mediu (scriptic) rezultă că el este un indicator de potențial, care reflectă disponibilul de forță de muncă din întreprindere, ce poate fi utilizat în procesul de producție.

Nivelul productivității muncii în acest caz este influențat de volumul absențelor de la lucru, indiferent de cauzele care le-au generat. Pentru înlăturarea acestor neajunsuri, nivelul productivității muncii poate fi calculat pe baza numărului mediu efectiv de personal.

Productivitatea muncii, determinată pe baza numărului de personal, reflectă producția medie realizată de o persoană în unitatea de timp şi poate fi: lunară, trimestrială, semestrială sau anuală (în funcție de perioada luată în calculul numărului mediu de personal).

Pentru calculul productivității muncii lunare se vor utiliza relațiile:

$$w_{ti} = \frac{Q_i}{T_i} \tag{10.48}$$

respectiv:

$$\overline{w}_t = \frac{\sum Q_i}{\sum T_i} = \frac{\sum w_{ti} T_i}{\sum T_i} = \sum w_{ti} y^{T_i} = \sum x_i^{\overline{w}_i} \tag{10.49}$$

în care: **i = 1, 2,…,n** sunt unități; $\mathbf{w_{ti}}$ – productivitatea medie lunară a muncii la unitatea **i**; $\mathbf{Q_i}$ – volumul producției la unitatea **i** (poate fi exprimat în unități naturale, natural convenționale, valorice sau unități de timp de muncă); $\mathbf{T_i}$ – număr mediu de personal (sau număr mediu efectiv) la unitatea **i**; $\overline{w}_t$ – productivitatea medie lunară calculată la nivelul unui grup de unități; y^{T_i} - structura numărului mediu de personal (sau a numărului mediu efectiv), pe unități; $x_i^{\overline{w}}$ - mărimea cu care contribuie fiecare unitate la formarea nivelului mediu al productivității muncii.

Factorii ce influențează variația productivității medii lunare sunt: productivitatea medie lunară calculată la nivelul unităților ($\mathbf{w_{ti}}$) şi structura numărului mediu de personal (respectiv a numărului mediu efectiv, pe întreprinderi y^{T_i}).

Analiza modificării în timp sau față de plan a productivității medii lunare se realizează cu ajutorul indicilor calculați ca raport a două medii (cu structură variabilă, fixă şi ai variației structurii).

Comparând dinamica productivității muncii calculată în funcție de numărul mediu de personal ($\mathbf{w/T_s}$) cu dinamica productivității muncii determinată pe baza numărului mediu efectiv (**w/**Tef) rezultă că prima este întotdeauna inferioară celeilalte, datorită proprietăților diferite în care este folosit numărul mediu de personal.

Material cu drept de autor

Între cele două productivități există o relație de forma:

$$w / T_s = w / T_{ef} . z \qquad (10.50)$$

în care **z = (T_{ef}/T_s)** este raportul ce se stabileşte între numărul mediu efectiv şi numărul mediu (scriptic).

În funcție de scopul analizei, productivitatea lunară a muncii poate fi calculată şi pe categorii de personal.

Legătura dintre productivitatea muncii determină la nivelul unei categorii de personal şi cea determinată pentru total personal se poate realiza prin intermediul unor coeficienți de recalculare, determinați ca raport între numărul personalului aferent categoriei pentru care s-a calculat nivelul productivității muncii şi numărul personalului aferent altei categorii.

Pentru determinarea coeficienților se utilizează structura personalului în funcție de locul pe care-l ocupă în procesul de producție, respectiv:

- muncitori, din care:
  - direct productivi:
  - indirect productivi;
- alte categorii de personal.

Productivitatea muncii lunare exprimată prin producția ce revine pe o persoană (w/T), poate fi prezentată în funcție de:

- productivitatea muncii muncitorilor direct productivi (**w/MD**);
- ponderea muncitorilor direct productivi în total muncitori (**K'**);
- ponderea muncitorilor în total personal (**K''**).

Deci:

$$w_{/T} = w_{/MD} K'K'' \qquad (10.51)$$

Dinamica productivității muncii lunare în acest caz se va determina:

$$I_{w/T} = \frac{w_{/MD(1)} K'_1 K''_1}{w_{/MD(0)} K'_0 K''_0} \qquad (10.52)$$

Influența separată a factorilor se poate determina cu ajutorul indicilor:

$$I_{w/T}^{w/MD} \; ; \; I_{w/T}^{K'} \; ; \; I_{w/T}^{K''}$$

Modificarea absolută a nivelului productivității muncii în acest caz se va calcula cu expresia:

$$\Delta_{w/T} = \Delta_{w/T}^{w/MD} + \Delta_{w/T}^{K'} + \Delta_{w/T}^{K''} \qquad (10.53)$$

Pentru a exemplifica calculul indicatorilor prezentați se consideră cunoscute următoarele date din activitatea unui agent economic pentru o lună din doi ani consecutivi, tabelul 10.6.

Pe baza datelor din tabelul 10.6 se poate analiza influența factorilor asupra variației productivității muncii calculată pe muncitor, respectiv a

Material cu drept de autor

productivității muncii calculată pe total personal. Dinamica productivității muncii pentru muncitori va fi:

$$I_{w/M} = \frac{w_{/M1}}{w_{/MD}} = \frac{w_{/MD1}K'_1}{w_{/MD0}K'_0} = \frac{6,944\times10^6\times0,7826}{6,667\times10^6\times0,75} =$$

$$= \frac{5,4345\times10^6}{5\times10^6} = 1,0869 \quad sau \quad 108,69\%$$

Tabelul 10.6

Perioada	Producția marfă în prețuri comparabile (mil. lei)	Număr mediu de personal				Productivitatea muncii (mil. lei)			Coeficienți corectori	
		Total	din care			Muncit. Direct prod.	Muncitori	Total personal	K'	K''
			muncitori		alte categ.					
			Direct product.	Ind. prod.						
de bază	720	160	108	36	16	6,667	5	4,5	0,75	0,9
curentă	875	175	126	39	14	6,944	5,435	5,0	0,7826	0,92

$$I_{w/M}^{w/MD} = \frac{w_{/MD1}K'_1}{w_{/MD0}K'_1} = \frac{5,4345\times10^6}{5,2176\times10^6} = 1,0416 \quad sau \quad 104,16\%$$

$$I_{w/M}^{K} = \frac{w_{/MD0}K'_1}{w_{/MD0}K'_0} = \frac{5,2176\times10^6}{5\times10^6} = 1,0435 \quad sau \quad 104,35\%$$

$$\Delta_{w/M} = w_{/M1} - w_{/M0} = (5,435-5)\times10^6 = 435\,mii\,lei/muncitor$$

$$\Delta_{w/M}^{K'} = w_{/MD0}(K'_1 - K'_0) = 6,67\times10^3(0,7826-0,75) =$$

$$= 217,344\,mii\,lei/muncitor$$

Se constată că ambii factori au avut o influență pozitivă asupra dinamicii productivității muncii calculate pe baza numărului de muncitori.

Folosind acelaşi sistem se poate descompune pe factori şi variația productivității muncii calculată în funcție de numărul total de personal.

Nivelul productivității muncii determinat pe baza numărului de personal are un caracter orientativ, întrucât nu ia în calcul timpul lucrat suplimentar.

În analiză se utilizează şi nivelul productivității muncii determinat pe baza cheltuielilor de timp de muncă exprimate în om-zile, ca productivitate zilnică:

$$w_{zi} = Q_i / T_{zi} \qquad (10.54)$$

respectiv:

Material cu drept de autor

$$\overline{w}_{zi} = \frac{\sum Q_i}{\sum T_{zi}} = \frac{\sum w_{zi} T_{zi}}{\sum T_{zi}} = \sum w_{zi} y^{T_{zi}} = \sum x_i^{\overline{w}_{zi}} \qquad (10.55)$$

în care: $\mathbf{w_{zi}}$ este productivitatea medie zilnică determinată la nivelul unităților; $\mathbf{T_{zi}}$ – timpul de muncă cheltuit exprimat în am-zile (se ia în calcul timpul lucrat în program normal de lucru şi timpul lucrat suplimentar); $y^{T_{zi}}$ - structura timpului de muncă cheltuit exprimat în om-zile; $x_i^{\overline{w}_{zi}}$ - mărimea cu care contribuie fiecare unitate la formarea nivelului productivității medii zilnice.

Factorii ce influențează dinamica productivității medii zilnice sunt: $\mathbf{w_{zi}}$ şi $y^{T_{zi}}$.

Pentru analiza variației în timp sau față de plan a productivității medii zilnice se calculează:

$$I_{\overline{w}_{zi}} \; ; \; I_{\overline{w}_{zi}}^{w_{zi}} \; ; \; I_{\overline{w}_{zi}}^{y^{T_{zi}}}$$

şi modificările absolute corespunzătoare.

Ca indicatori de efect economic se calculează plusul sau minusul de producție datorat variației productivității medii zilnice în general şi pe factori astfel:

$$\Delta_{\sum Q}^{w_{zi}} = \Delta_{\overline{w}_{zi}} \times \sum T_{zi1} \qquad (10.56)$$

$$\Delta_{\sum Q}^{w_{zi}} = \left(\sum w_{zi1} y_1^{T_{zi}} - \sum w_{zi0} y_1^{T_{zi}}\right) \times T_{zi1} \qquad (10.57)$$

$$\Delta_{\sum Q}^{y^{T_{zi}}} = \left(\sum w_{zi0} y_1^{T_{zi}} - \sum w_{zi0} y_0^{T_{zi}}\right) \times T_{zi1} = \Delta_{\overline{w}_{zi}}^{y^{T_{zi}}} \times \sum T_{zi1} \qquad (10.58)$$

Un grup de întreprinderi, situația este bună în cazul în care $I_{\overline{w}_{zi}} > 1$, respectiv $\Delta_{\overline{w}_{zi}} > 0$. Această creştere trebuie să se realizeze cu prioritate pe seama variației factorului calitativ ($\mathbf{w_{zi}}$).

Productivitatea muncii, prin definiție, este un indicator ce măsoară eficiența muncii vii şi impune folosirea în cuantificare a nivelului ei, a cheltuielilor de muncă efective, care au contribuit direct sau indirect la realizarea producției.

Aceste cheltuieli sunt exprimate în om-ore, iar pe baza lor se calculează productivitatea orară a muncii:

$$w_{h_i} = \frac{Q_i}{T_{hi}} \qquad (10.59)$$

$$\overline{w}_h = \frac{\sum Q_i}{\sum T_{hi}} = \frac{\sum w_{hi} \times T_{hi}}{\sum T_{hi}} = \sum w_{hi} \times y^{T_{hi}} = \sum x_i^{\overline{w}_h} \qquad (10.60)$$

Material cu drept de autor

în care w_{hi} este productivitatea medie orară a muncii în unitatea **i**; Q_i – volumul producției la unitatea **i**; T_{hi} – timp de muncă exprimat în om-ore (se ia în calcul timpul lucrat în regim normal de lucru şi timpul lucrat suplimentar); $y^{T_{hi}}$ - structura timpului de muncă exprimat în om-ore; $x_i^{\overline{w_h}}$ - mărimea cu care contribuie fiecare unitate la formarea nivelului mediu a productivității muncii orare, calculată la nivelul unui grup de unități.

Factorii ce influențează dinamica productivității muncii orare sunt w_{hi} şi $y^{T_{hi}}$. Dinamica acesteia se va calcula cu ajutorul indicatorilor:

$$I_{\overline{w}_h} \; ; \; I_{\overline{w}_h}^{w_{hi}} \; ; \; I_{\overline{w}_h}^{y^{T_{hi}}}$$

respectiv:

$$\Delta_{\overline{w}_h} \; ; \; \Delta_{\overline{w}_h}^{w_{hi}} \; ; \; \Delta_{\overline{w}_h}^{y^{T_{hi}}}$$

Modificarea absolută a volumului producției datorate variației productivității medii orare în general şi pe factori este dată de relațiile:

$$\Delta_{\sum Q}^{\overline{w}_h} = \Delta_{\overline{w}_h} \times \sum T_{hi1} \qquad (10.61)$$

$$\Delta_{\sum Q}^{w_{hi}} = \left(\sum w_{hi1} y_1^{T_{zi}} - \sum w_{hi0} y_1^{T_{zi}}\right) \times T_{hi1} = \Delta_{\overline{w}_h}^{w_{hi}} T_{hi1} \qquad (10.62)$$

$$\Delta_{\sum Q}^{y^{T_{hi}}} = \left(\sum w_{hi0} y_1^{T_{hi}} - \sum w_{hi0} y_0^{T_{hi}}\right) \times T_{hi1} = \Delta_{\overline{w}_{hi}}^{y^{T_{hi}}} \times \sum T_{hi1} \qquad (10.63)$$

Creşterea productivității medii orare trebuie să se realizeze în principal pe seama sporiri productivității medii orare individuale (w_{hi}). În nivelul acesteia sunt sintetizate influențele unui complex de factori: gradul de înzestrare a producției şi a muncii, nivelul calificării muncitorilor etc.

Identificarea rezervelor de creştere a productivității muncii orare este condiționată de calculul acesteia pe categorii de personal şi în special pentru muncitorii direct productivi:

$$w_{h/MD} = Q / T_{h/MD} \qquad (10.64)$$

în care $w_{h/MD}$ este productivitatea orară a muncitorilor direct productivi; **Q** – producția realizată; $T_{h/MD}$ – timp de muncă cheltuit de muncitorii direct productivi exprimată în ore-om.

Ştiind că volumul producției poate fi calculat:

$$Q = w_{h/MD} \times T_{h/M} \qquad (10.65)$$

respectiv:

$$Q = r_h \times T_{hu} \qquad (10.66)$$

Nivelul productivității muncii orare este dat de relația:

$$w_{h/MD} = r_h \times \left(T_{hu} / T_{h/MD}\right) \qquad (10.67)$$

Material cu drept de autor

în care $\mathbf{r_h}$ este randamentul orar al utilajelor (volumul fizic de producție realizat într-o oră); $\mathbf{T_{hu}}$ – timpul de funcționare al utilajelor exprimat în maşini-ore.

Analizând dinamica productivității muncii orare pe baza relației:

$$w_{h/MD} = r_h \times \left(T_{hu} / T_{h/MD}\right)$$

se constată că modificarea acesteia este influențată de variația randamentului orar al utilajelor şi de raportul ce se stabileşte între timpul de funcționare al utilajelor şi de timpul efectiv lucrat de către muncitorii direct productivi.

10.2. Analiza statistică a dinamicii fondului de salarii

Fondul de salarii este un indicator a cărui dinamică este influențată de variația salariului mediu şi a numărului de personal.

Analiza modificării în timp, în general şi pe factori a fondului de salarii se realizează cu ajutorul unui sistem de indicatori ce pot fi determinați la nivelul unităților componente ale sistemului (ca indicatori individuali), sau ca indicatori de grup, la nivelul sistemului.

a. La nivelul unităților componente ale sistemului se calculează:

- **indicele dinamicii fondului de salarii (i_{si}) este**:

$$i_{si} = \frac{s_{i1}}{s_{i0}} \qquad (10.68)$$

- **modificarea relativă a fondului de salarii în perioada curentă față de perioada de bază (R_s)**:

$$R_{si} = i_{si} - 1 = \frac{\Delta_{si}}{s_{i0}} \qquad (10.69)$$

- **modificarea absolută a fondului de salarii în perioada curentă față de perioada de bază (Δ_{si})**:

$$\Delta_{si} = s_{i1} - s_{i0} = R_{si} \times s_{i0} \qquad (10.70)$$

în care: $\mathbf{s_{i0}}$; $\mathbf{s_{i1}}$ este fondul de salarii la nivelul unităților componente ale sistemului în perioada de bază, respectiv curentă; **i = 1, 2,...,n** – unități componente ale sistemului (categorii de personal, secții, ateliere etc.).

Fondul de salarii la acest nivel poate fi calculat ca un produs între salariul mediu ($\mathbf{s_i}$) şi numărul de salariați ($\mathbf{T_i}$), respectiv:

$$S_i = s_i \times T_i \qquad (10.71)$$

Deci, modificarea în timp a fondului de salarii este influențată de variația celor doi factori:

Material cu drept de autor

$$i_{si} = \frac{S_{i1}}{S_{i0}} = \frac{s_{i1}T_{i1}}{s_{i0}T_{i0}} = i_{s_i} \times i_{T_i} \qquad (10.72)$$

Dinamica fondului de salarii este influențată de dinamica salariului mediu şi a numărului de salariați.

Influența factorilor asupra modificării absolute a fondului de salarii se calculează cu ajutorul relațiilor:

$$\Delta_{si} = S_{i1} \times S_{i0} = \begin{cases} \Delta_{s_i}^{s_i} = \Delta_{s_i} \times T_{i1} \\ \Delta_{s_i}^{T_i} = s_{i0} \times \Delta_{T_i} \end{cases} \qquad (10.73)$$

În cazul în care modificările absolute datorate influenței separate a factorilor au acelaşi semn, în analiza statistică este util să se determine cât din modificarea absolută a fondului de salarii revine fiecărui factor:

$$K_1 = \frac{\Delta_{S_i}^{s_i}}{\Delta_{s_i}} \text{ respectiv, } K_2 = \frac{\Delta_{S_i}^{T_i}}{\Delta_{s_i}}, \; K_1 + K_2 = 1$$

b. La nivelul sistemului (întreprindere, grup de întreprinderi etc.), indicatorii dinamici ai fondului de salarii se calculează:

$$I_{\sum s} = \frac{\sum S_1}{\sum S_0} \qquad (10.74)$$

$$R_{\sum s} = I_{\sum s} - 1 = \frac{\Delta_{\sum s}}{\sum S_0} \qquad (10.75)$$

$$\Delta_{\sum s} = \sum S_1 - \sum S_0 = R_{\sum s} \times \sum S_0 \qquad (10.76)$$

Pentru evidențierea influenței factorilor asupra variației fondului de salarii se utilizează sistemele de indici calculați sub formă agregată, constituiți în două variante:

1. în cazul în care factorul structural este cuprins în factorul cantitativ, dinamica fondului total de salarii se calculează:

$$I_{\Sigma S_i} = \frac{\sum S_{i1}}{\sum S_{i0}} = \frac{\sum s_{i1} \times T_{i1}}{\sum s_{i0} \times T_{i0}} \qquad (10.77)$$

iar modificarea absolută este dată de relația:

$$\Delta_{\sum s_i} = \sum S_{i1} - \sum S_{i0} = \sum s_{i1}T_{i1} - \sum s_{i0}T_{i0} = \sum \Delta_{s_i} \qquad (10.78)$$

Factorii ce influențează dinamica fondului de salarii în acest caz sunt:

Material cu drept de autor

- modificarea salariului mediu pe categorii de personal ($\mathbf{s_i}$);
- modificarea numărului de personal pe categorii ($\mathbf{T_i}$).

Indicii factoriali sunt:

$$I_{\Sigma s_i}^{s_i} = \frac{\sum s_{i1} \times T_{i1}}{\sum s_{i0} \times T_{i1}} \qquad (10.79)$$

şi:

$$I_{\Sigma s_i}^{T_i} = \frac{\sum s_{i0} \times T_{i1}}{\sum s_{i0} \times T_{i0}} \qquad (10.80)$$

Modificarea absolută a fondului de salarii ce evidenţiază influenţa separată a factorilor se va calcula:

$$\Delta_{\Sigma S_i}^{s_i} = \Sigma s_{i1} T_{i1} - \Sigma s_{i0} T_{i1} = \Sigma \Delta_{s_i} \times T_{i1} = \Sigma \Delta_{S_i}^{s_i} \qquad (10.81)$$

$$\Delta_{\Sigma S_i}^{T_i} = \Sigma s_{i0} T_{i1} - \Sigma s_{i0} T_{i0} = \Sigma \Delta_{s_{i0}} \times \Delta_{T_i} = \Sigma \Delta_{S_i}^{T_i} \qquad (10.82)$$

Relaţiile de verificare sunt:

$$I_{\Sigma S_i} = I_{\Sigma S_i}^{s_i} \times I_{\Sigma S_i}^{T_i} \qquad (10.83)$$

$$\Delta_{\Sigma S_i} = \Delta_{\Sigma S_i}^{s_i} + \Delta_{\Sigma S_i}^{T_i} \qquad (10.84)$$

2. în cazul în care factorul structural este cuprins în factorul calitativ, dinamica fondului total de salarii se calculează cu ajutorul relaţiei:

$$I_{\Sigma S_i} = \frac{\sum S_{i1}}{\sum S_{i0}} = \frac{\bar{s}_1 \sum T_{i1}}{\bar{s}_0 \sum T_{i0}} \qquad (10.85)$$

iar modificarea absolută:

$$\Delta_{\Sigma s_i} = \Sigma S_{i1} - \Sigma S_{i0} = \bar{s}_1 \Sigma T_{i1} - \bar{s}_0 \Sigma T_{i0} \qquad (10.86)$$

Dinamica fondului de salarii în acest caz este influenţat de variaţia:

- salariului mediu ($\bar{s}$) calculat la nivel de întreprindere (ce cuprinde şi influenţa factorului structural):
- numărul total de personal (ΣT_i).

Indicii factoriali sunt:

$$I_{\Sigma S_i}^{\bar{s}} = \frac{\bar{s}_1 \sum T_{i1}}{\bar{s}_0 \sum T_{i1}} = I_{\bar{s}} \qquad (10.87)$$

şi:

Material cu drept de autor

$$I_{\Sigma S_i}^{\Sigma T_i} = \frac{\bar{s}_0 \sum T_{i1}}{\bar{s}_0 \sum T_{i0}} = I_{\Sigma T_i} \qquad (10.88)$$

Modificarea absolută a fondului total de salarii ce evidențiază influența separată a factorilor se calculează:

$$\Delta_{\Sigma S_i}^{\bar{s}} = \bar{s}_1 \Sigma T_{i1} - \bar{s}_0 \Sigma T_{i0} = \Delta_{\bar{s}} \times \Sigma T_{i1} \qquad (10.89)$$

$$\Delta_{\Sigma S_i}^{\Sigma T_i} = \bar{s}_0 \Sigma T_{i1} - \bar{s}_0 \Sigma T_{i0} = \bar{s}_0 \Delta_{\Sigma T_i} \qquad (10.90)$$

Relațiile de verificare sunt:

$$I_{\Sigma S_i} = I_{\Sigma S_i}^{\bar{s}} \times I_{\Sigma S_i}^{\Sigma T_i} = I_{\bar{s}} \times I_{\Sigma T_i} \qquad (10.91)$$

$$\Delta_{\Sigma S_i} = \Delta_{\Sigma S_i}^{\bar{s}} + \Delta_{\Sigma S_i}^{\Sigma T_i} \qquad (10.92)$$

Cele două cazuri se folosesc la analiza în dinamică a fondului de salarii în funcție de necesitatea localizării acțiunii factorilor de influență. În primul caz, factorii de influență sunt localizați la nivelul categoriei de personal, secții, ateliere etc., iar în cel de al doilea caz, factorii de influență sunt localizați la nivel de întreprindere.

Pentru calculul indicatorilor prezentați se consideră cunoscute următoarele date din activitatea unei întreprinderi pentru lunile martie şi aprilie (indicatorii au fost calculați pentru acelaşi număr de zile lucrătoare).

Tabelul 10.7

Secții	Fondul de salarii (mil. lei)		Număr de muncitori		Salariul mediu (mii lei/muncitor)		$S_{i0} \times T_{i1}$, mil.lei
	0	1	0	1	0	1	
1	62	108,68	100	143	620	760	88,66
2	84	74,88	150	117	560	640	65,52
Total	146	183,56	250	260	584	706	154,18

Se va exemplifica calculul dinamicii fondului de salarii pentru secția 1. Indicii şi modificările absolute sunt:

$$i_s = \frac{S_1}{S_0} = \frac{108,68 \times 10^6}{62 \times 10^6} = 1,7529 \; sau \; 175,29\%$$

$$\Delta_s = S_1 - S_0 = 108,68 \times 10^6 - 62 \times 10^6 = 46,68 mil.lei$$

$$i_s^s = \frac{\bar{S}_1}{\bar{S}_0} = \frac{760 \times 10^3 \times 143}{620 \times 10^3 \times 143} = 1,2258 \; sau \; 122,58\%$$

Material cu drept de autor

$$\Delta_s^s = \overline{S}_1 - \overline{S}_0 = (760-620)\times 10^3 \times 143 = 20{,}02\, mil.lei$$

$$i_s^T = \frac{\overline{S}_1}{\overline{S}_0} = \frac{620\times 10^3 \times 143}{620\times 10^3 \times 100} = 1{,}43\ sau\ 143\%$$

$$\Delta_s^T = 620\times 10^3 \times (143-100) = 26{,}66\, mil.lei$$

Pentru secția 1 s-a înregistrat o creştere a fondului de salarii cu 75,29%, respectiv 46,68 mil. lei, din care 57,155% s-a datorat creşterii numărului mediu de muncitori.

Pentru secția 2-a indicatorii s-au calculat la fel, iar rezultatele sunt prezentate în tabelul 10.8.

Tabelul 10.8

Secții	% i_{Si}	Mil.lei Δ_{Si}	% $i_{s_i}^{s_i} = i_{S_i}$	Mil.lei $\Delta_{s_i}^{s_i}$	% $i_{s_i}^{T_i} = i_{T_i}$	Mil. lei $\Delta_{s_i}^{T_i}$
1	175,29	46,68	122,58	20,02	143	26,66
2	89,14	-9,12	114,29	9,36	78	-18,48
Total	125,73	37,56	119,06	29,38	105,60	8,18

La nivelul întreprinderii, situația se prezintă astfel:

1.

$$I_{\Sigma S_i} = \frac{\sum S_{i1}}{\sum S_{i0}} = \frac{\sum s_{i1}\times T_{i1}}{\sum s_{i0}\times T_{i0}} = \frac{183{,}56\times 10^6}{146\times 10^6} = 1{,}2573\ sau\ 125{,}73\%$$

$$\Delta_{\sum s_i} = \sum S_{i1} - \sum S_{i0} = \sum s_{i1}T_{i1} - \sum s_{i0}T_{i0} = \sum \Delta_{s_i}$$

$$= (183{,}56-146)\times 10^6 = 37{,}56\, mil.\ lei$$

$$I_{\Sigma s_i}^{s_i} = \frac{\sum s_{i1}\times T_{i1}}{\sum s_{i0}\times T_{i1}} = \frac{183{,}56\times 10^6}{154{,}18\times 10^6} = 1{,}1906\ sau\ 119{,}06\%$$

$$I_{\Sigma s_i}^{T_i} = \frac{\sum s_{i0}\times T_{i1}}{\sum s_{i0}\times T_{i0}} = \frac{154{,}18\times 10^6}{146\times 10^6} = 1{,}056\ sau\ 105{,}6\%$$

$$\Delta_{\Sigma S_i}^{s_i} = \Sigma s_{i1}T_{i1} - \Sigma s_{i0}T_{i1} = \Sigma \Delta_{s_i}\times T_{i1} = \Sigma \Delta_{S_i}^{s_i} =$$

$$= (183{,}56-154{,}18)\times 10^6 = 29{,}38\ mil.\ lei$$

Material cu drept de autor

$$\Delta_{\Sigma S_i}^{T_i} = \Sigma s_{i0} T_{i1} - \Sigma s_{i0} T_{i0} = \Sigma \Delta_{s_{i0}} \times \Delta_{T_i} = \Sigma \Delta_{S_i}^{T_i} =$$

$$= (154{,}18 - 146) \times 10^6 = 8{,}18 \; mil. \; lei$$

În acest caz, modificarea totală a fondului de salarii de 40,2 mil. lei s-a repartizat pe fiecare astfel:

- pe seama variației salariului mediu pe secții 32,02 mil. lei:

$$32{,}02 \times 10^6 = (20{,}02 + 12) \times 10^6$$

- pe seama variației numărului de muncitori pe secții 8,18 mil. lei:

$$8{,}18 \times 10^6 = (26{,}66 - 18{,}48) \times 10^6$$

2.

$$I_{\Sigma S_i} = \frac{\sum S_{i1}}{\sum S_{i0}} = \frac{183{,}56 \times 10^6}{146 \times 10^6} = 1{,}2573 = \begin{cases} I_{\Sigma S}^{\bar{s}} = 1{,}2089 \\ I_{\Sigma S}^{\Sigma T} = 1{,}04 \end{cases}$$

$$\Delta_{\Sigma s_i} = \Sigma S_{i1} - \Sigma S_{i0} = 37{,}56 = \begin{cases} \Delta_{\Sigma S}^{\bar{s}} = 31{,}72 mil.lei \\ \Delta_{\Sigma S}^{\Sigma T} = 5{,}84 mil. \; lei \end{cases}$$

În acest caz, 84,45% din modificarea totală a fondului de salarii s-a datorat creşterii salariului mediu cu 20,89%.

10.3. Analiza statistică a dinamicii salariului mediu

Salariul mediu se calculează pe categorii de personal, iar pentru muncitori se poate determina în funcție de gruparea acestora supă o serie de caracteristici.

La nivelul categoriei de personal, dinamica salariului mediu se determină astfel:

$$i_s = \frac{s_{i1}}{s_{i0}} = \frac{S_{i1}}{T_{i1}} : \frac{S_{i0}}{T_{i0}} = \frac{S_{i1}}{S_{i0}} : \frac{T_{i1}}{T_{i0}} = i_s : i_T \qquad (10.93)$$

Deci, indicele salariului mediu pentru categoria **i** poate fi calculat ca un raport între indicele fondului de salarii aferent categoriei **i** şi indicele numărului de personal din categoria respectivă.

Dinamica salariului salarului mediu la nivel de întreprindere se calculează cu ajutorul indicilor de grup cu baza de medie astfel:

$$I_{\bar{s}} = \frac{\bar{S}_1}{\bar{S}_0} = \frac{\Sigma S_{i1}}{\Sigma T_{i1}} : \frac{\Sigma S_{i0}}{\Sigma T_{i0}} = \frac{\Sigma S_{i1}}{\Sigma S_{i0}} : \frac{\Sigma T_{i1}}{\Sigma T_{i0}} = I_{\Sigma S_i} : I_{\Sigma T_i} \qquad (10.94)$$

Material cu drept de autor

sau:

$$I_{\bar{s}} = \frac{\bar{S}_1}{\bar{S}_0} = \frac{\Sigma s_{i1} T_{i1}}{\Sigma T_{i1}} : \frac{\Sigma s_{i0} T_{i0}}{\Sigma T_{i0}} = \frac{\Sigma s_{i1} y_1^{T_i}}{\Sigma s_{i0} y_0^{T_i}} = \frac{\Sigma x_{i1}^{\bar{s}}}{\Sigma x_{i0}^{\bar{s}}} \qquad (10.95)$$

în care $\bar{s}$ este salariul mediu; **i = 1, 2,...,n** – categorii de personal (secții, ateliere etc.); y^{T_i} - structura personalului pe categorii; $x_i^{\bar{s}}$ - mărimea cu care contribuie categoria de salariați **i** la formarea salarului mediu.

Deci, dinamica salarului mediu la nivel de întreprindere este influențată de variația:

- salarului mediu pe categorii de personal, secții etc.;
- structura salariaților.

Modificarea absolută a salarului mediu se calculează:

$$\Delta_{\bar{s}} = \bar{s}_1 - \bar{s}_0 = \Sigma s_{i1} y_1^{T_i} - \Sigma s_{i0} y_0^{T_i} \qquad (10.96)$$

sau:

$$\Delta_{\Sigma s}^{\bar{s}} = \Delta_s \times \Sigma T_{i1} \qquad (10.97)$$

Influența separată a factorilor este dată de $I_{\bar{s}}^{s_i}$ şi $I_{\bar{s}}^{y^{T_i}}$, $\Delta_{\bar{s}}^{s_i}$ şi $\Delta_{\bar{s}}^{y^{T_i}}$,respectiv: $\Delta_{\Sigma S_i}^{s_i}$ şi $\Delta_{\Sigma S_i}^{y^{T_i}}$.

Pentru exemplificare privind calculul salarului mediu se vor relua datele din exemplul anterior, tabelul 10.9.

Tabelul 10.9

secții	s_i, (mii lei/persoană)		y^{T_i} (%)		$x_i^{\bar{s}} = s_i \times y^{T_i}$ (mii lei/persoană)		
	0	1	0	1	0	1	*
1	620	760	40	55	248	418	341
2	560	640	60	45	336	288	252
Total	584	706	100	100	584	706	593

$$I_{\bar{s}} = \frac{\bar{S}_1}{\bar{S}_0} = \frac{\Sigma x_{i1}^{\bar{s}}}{\Sigma x_{i0}^{\bar{s}}} = \frac{706 \times 10^3}{584 \times 10^3} = 1{,}2089 \ \ sau \ 120{,}89\%$$

$$\Delta_{\bar{s}} = \bar{s}_1 - \bar{s}_0 = (706 - 584) \times 10^3 = 122 \, mii \, lei / pers.$$

$$\Delta_{\Sigma s}^{\bar{s}} = \Delta_s \times \Sigma T_{i1} = (706 - 584) \times 10^3 \times 260 = 31{,}72 \, mil.lei$$

Material cu drept de autor

$$I_{\bar{s}}^{s_i} = \frac{\bar{S}_1}{\bar{S}_*} = \frac{\Sigma x_{i1}^{\bar{s}}}{\Sigma x_{i*}^{\bar{s}}} = \frac{706 \times 10^3}{593 \times 10^3} = 1{,}19055 \ \ sau \ 119{,}055\%$$

$$\Delta_{\bar{s}}^{s_i} = \bar{s}_1 - \bar{s}_* = (706 - 593) \times 10^3 = 113 \, mii \, lei / pers.$$

$$\Delta_{\Sigma s}^{\bar{s}} = (706 - 593) \times 10^3 \times 260 = 29{,}38 \, mil.lei$$

$$I_{\bar{s}}^{y^{T_i}} = \frac{593 \times 10^3}{584 \times 10^3} = 1{,}0154 \ \ sau \ 101{,}54\%$$

$$\Delta_{\bar{s}}^{y^{T_i}} = (593 - 584) \times 10^3 = 9 \, miloane \ lei / persoană$$

$$\Delta_{\Sigma s}^{x^{T_i}} = (593 - 584) \times 10^3 \times 260 = 2{,}34 \, mil.lei$$

Se constată că modificarea salarului mediu pe secții a influențat în sensul creşterii fondul de salarii pe întreprindere cu 29,38 milioane lei.

Schema influenței factorilor asupra modificării absolute a fondului de salarii se prezintă astfel:

$$\Delta_{\Sigma S_i} = 37{,}56 \, mil.lei = \begin{cases} a) \ \Delta_{\Sigma s_i} = \begin{cases} \Delta_{\Sigma s_i}^{s_i} = 29{,}38 \, mil.lei \\ \Delta_{\Sigma s_i}^{T_i} = 8{,}18 \, mil.lei \end{cases} \\ b) \ \Delta_{\Sigma s_i}^{\bar{s}} = \begin{cases} 31{,}72 \, mil.lei \begin{cases} \Delta_{\Sigma s_i}^{s_i} = 29{,}38 \, mil.lei \\ \Delta_{\Sigma s_i}^{y^{T_i}} = 2{,}34 \, mil.lei \end{cases} \\ \Delta_{\Sigma s_i}^{\Sigma T_i} = 5{,}84 \, mil.lei \end{cases} \end{cases}$$

Exemplu de calcul 10.1. Pentru trei societăți comerciale se cunosc datele statistice din tabelul 10.10:

Tabelul 10.10. Date inițiale

S.C.	Valoarea producției în prețuri curente (u.m.)	Modificarea prețurilor (%)	Dinamica volumului fizic al producției (%)	Număr salariați	
				PB	PC
0	1	2	3	4	5
A	500	+50	80	50	45
B	1200	+80	100	150	150
C	300	+100	90	60	58

Material cu drept de autor

Se cere:

1. Indicii individuali ai preţurilor şi ai valorii.

2. Indicii de grup ai valorii, volumului fizic şi al preţurilor.

3. Modificarea relativă şi absolută a valorii producţiei, cu evidenţierea influenţei factorilor.

4. Indicii individuali şi de grup ai numărului de salariaţi.

5. Productivitatea muncii, pe fiecare societate comercială şi pe total, în cele două perioade: de Bază (PB) şi curentă (PC).

6. Indicii individuali şi de grup ai productivităţii muncii.

7. Modificarea relativă şi absolută a productivităţii medii a muncii cu evidenţierea factorilor de influenţă.

8. Modificarea producţiei pe seama productivităţii şi a numărului de salariaţi, cu descompunerea pe doi factori şi pe trei factori de influenţă.

9. Să se reprezinte grafic producţia pe total ţinând seama de factorii de influenţă (pentru cazul a doi factori de influenţă).

Rezolvare.

1. Metodologia şi rezultatele calculelor, pe baza datelor iniţiale, rezultă din tabelul 10.11, coloanele 3 şi 5.

Tabelul 10.11. Calcule intermediare.

SC	p_1q_1 (u.m.)	$r_{1/0}^{p}$	$i_{1/0}^{p} = r_{1/0}^{p} + 100$ (%)	$i_{1/0}^{q}$
0	1	2	3	4
A	500	+50	150	80
B	1200	+80	180	100
C	300	+100	200	90
Total	2000	-	-	-

Tabelul 10.11 (continuare)

$i_{1/0}^{v} = i_{1/0}^{p} \times i_{1/0}^{q}$	$i_{1/0}^{v}$ %	$p_0q_0 = \frac{p_1q_1}{i_{1/0}^{v}}$	$p_0q_1 = \frac{p_1q_1}{i_{1/0}^{p}}$
5	6	7	8
1,2	120	416,667	333,333
1,8	180	666,667	666,667
1,8	180	166,667	150
-	-	1250	1150

Material cu drept de autor

2. Pentru calculul indicilor de grup trebuie obținute următoarele informații: valoarea producției în perioada de bază (q_0p_0), valoarea producției în prețuri constante (p_0q_1). Rezultatul calculelor sunt prezentate în tabelul 10.11, coloanele 1, 7 şi 8.

$$I_{1/0}^{v(p,q)} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} = \frac{2000}{1250} = 1{,}6 \Rightarrow 160\%$$

$$I_{1/0}^{v(q)} = \frac{\Sigma i_{1/0}^{q} p_0 q_0}{\Sigma p_0 q_0} = \frac{1150}{1250} = 0{,}92 \Rightarrow 92\%$$

$$I_{1/0}^{v(p)} = \frac{\Sigma p_1 q_1}{\Sigma \frac{1}{i_{1/0}^{p}} p_1 q_1} = \frac{2000}{\frac{500}{1{,}5} + \frac{1200}{1{,}8} + \frac{300}{2}} = \frac{2000}{1150} = 1{,}7391 \Rightarrow 173{,}91\%$$

3. Modificarea relativă şi absolută a valorii producției, cu evidențierea influenței factorilor.

• modificările relative:

► pentru factorul cantitativ: $R_{1/0}^{v(q)} = I_{1/0}^{v(q)}\% - 100 = -8\%$

► pentru factorul calitativ: $R_{1/0}^{v(p)} = I_{1/0}^{v(p)}\% - 100 = +73{,}91\%$

► pe total: $R_{1/0}^{v(p,q)} = I_{1/0}^{v(p,q)}\% - 100 = +60\%$

• modificările relative:

► pe seama factorului cantitativ:

$$\Delta_{1/0}^{v(q)} = \Sigma i_{1/0}^{q} p_0 q_0 - \Sigma p_0 q_0 = -100 \ u.m.$$

► pe seama factorului calitativ:

$$\Delta_{1/0}^{v(p)} = \Sigma p_1 q_1 - \Sigma \frac{i}{i_{1/0}^{p}} p_1 q_1 = 850 \ u.m.$$

► pe total: $\Delta_{1/0}^{v(p,q)} = \Sigma p_1 q_1 - \Sigma p_0 q_0 = 750 \ u.m.$

4. Indicii individuali şi de grup ai numărului de salariați se determină cu expresiile:

$$i_{1/0}^{T} = \frac{T_1}{T_0} = \begin{cases} SC"A" \Rightarrow \frac{45}{50} = 0{,}9 \Rightarrow 90\% \\ SC"B" \Rightarrow \frac{150}{150} = 1{,}0 \Rightarrow 100\% \\ SC"C" \Rightarrow \frac{58}{60} = 0{,}9667 \Rightarrow 96{,}67\% \end{cases}$$

Material cu drept de autor

$$I_{1/0}^{T}=\frac{\Sigma T_1}{\Sigma T_0}=\frac{253}{260}=0{,}9731\Rightarrow 97{,}31\%$$

5. Expresiile de calcul şi rezultatele calculelor pentru productivitatea muncii pe fiecare societate comercială sunt prezentate în tabelul 10.12, coloanele 5 şi 6.

Tabelul 10.12. Calcule intermediare.

SC	p_0q_0 (u.m.)	p_0q_1	T_0	T_1	$w_0=\frac{p_0q_0}{T_0}$
0	1	2	3	4	5
A	416,667	333,333	50	45	8,333
B	666,667	666,667	150	150	4,444
C	166,667	150	60	58	2,777
Total	**1250**	**1150**	**260**	**253**	**4,808**

Tabelul 10.12 (continuare)

$w_1=\frac{p_0q_1}{T_1}$	$i_{1/0}^{w}=\frac{w_1}{w_0}$	w_0T_1	$i_{1/0}^{T}$
6	7	8	9
7,407	0,8889	373,985	0,9
4,444	1,0	666,600	1,0
2,586	0,9312	161,067	0,9667
4,545	**0,9453**	**1202,652**	**0,9731**

6. Indicii individuali ai productivității muncii sunt prezentați în tabelul 10.12, coloana 7.

Pentru calculul indicilor de grup ai productivității medii a muncii trebuie calculate productivitățile medii pe total, astfel:

$$\overline{w}_0=\frac{\Sigma w_0T_0}{\Sigma T_0}=\frac{\Sigma p_0q_0}{\Sigma T_0}=4{,}808\ u.m./\ salariat$$

$$\overline{w}_1=\frac{\Sigma w_1T_1}{\Sigma T_1}=\frac{\Sigma p_0q_1}{\Sigma T_1}=4{,}545\ u.m./\ salariat$$

Material cu drept de autor

$$\overline{w}_* = \frac{\Sigma w_0 T_1}{\Sigma T_1} = \frac{1202{,}652}{253} = 4{,}753\ u.m./\ salariat$$

Pe baza acestora se calculează următorii indicatori:

► indicele cu structură variabilă:

$$I_{SV}^{\overline{w}(w,g^T)} = \frac{\overline{w}_1}{\overline{w}_0} = \frac{4{,}545}{4{,}808} = 0{,}9453 \Rightarrow 94{,}53\%$$

► indicele variației structurii:

$$I_{SV}^{\overline{w}(g^T)} = \frac{\overline{w}_*}{\overline{w}_0} = \frac{4{,}753}{4{,}808} = 0{,}9896 \Rightarrow 98{,}96\%$$

► indicele cu structură fixă:

$$I_{SV}^{\overline{w}(w)} = \frac{\overline{w}_1}{\overline{w}_*} = \frac{4{,}545}{4{,}753} = 0{,}9562 \Rightarrow 95{,}62\%$$

7. Pe baza indicilor de grup se pot calcula:

• modificările absolute:

$$\Delta_{1/0}^{\overline{w}(w,g^T)} = \overline{w}_1 - \overline{w}_0 = 4{,}545 - 4{,}808 = -0{,}263\ u.m.$$

$$\Delta_{1/0}^{\overline{w}(g^T)} = \overline{w}_* - \overline{w}_0 = 4{,}753 - 4{,}808 = -0{,}055\ u.m.$$

$$\Delta_{1/0}^{\overline{w}(w)} = \overline{w}_1 - \overline{w}_* = 4{,}545 - 4{,}753 = -0{,}208\ u.m.$$

• modificările relative:

$$R_{1/0}^{\overline{w}(w,g^T)} = -5{,}47\%;\ R_{1/0}^{\overline{w}(g^T)} = -1{,}14\%;\ R_{1/0}^{\overline{w}(w)} = -4{,}38\%.$$

8.a. Descompunerea producției pe doi factori de influență care se poate realiza în mai multe moduri:

► varianta I.

$$I^{Q(w,T)} = \frac{\Sigma w_1 T_1}{\Sigma w_0 T_0} = \frac{\Sigma p_0 q_1}{\Sigma p_0 q_0} = 0{,}92 \Rightarrow 92\% \Rightarrow$$

$$\Rightarrow R^{Q(w,T)} = -8\%; \quad \Delta^{Q(w,T)} = -100\ u.m.$$

$$I^{Q(T)} = \frac{\Sigma w_0 T_1}{\Sigma w_0 T_0} = \frac{1202{,}652}{1250} = 0{,}9621 \Rightarrow 96{,}21\% \Rightarrow$$

$$\Rightarrow R^{Q(T)} = -3{,}79\%; \quad \Delta^{Q(T)} = -47{,}348\ u.m.$$

Material cu drept de autor

$$I^{Q(w)} = \frac{\Sigma w_1 T_1}{\Sigma w_0 T_1} = \frac{1150}{1202,652} = 0,95621 \Rightarrow 95,62\% \Rightarrow$$

$$\Rightarrow R^{Q(w)} = -4,38\%; \quad \Delta^{Q(T)} = -52,652 \; u.m.$$

► varianta II.

$$I^{Q(\overline{w},\Sigma T)} = \frac{\overline{w}_1 \Sigma T_1}{\overline{w}_0 \Sigma T_0} = 0,92 \Rightarrow 92\% \Rightarrow$$

$$\Rightarrow R^{Q(\overline{w},\Sigma T)} = -8\%; \; \Delta^{Q(\overline{w}\Sigma T)} = -100 \; u.m.$$

$$I^{Q(\Sigma T)} = \frac{\overline{w}_0 \Sigma T_1}{\overline{w}_0 \Sigma T_0} = 0,9731 \Rightarrow 97,31\% \Rightarrow R^{Q(\Sigma T)} = -2,69\%$$

$$\Delta^{Q(\Sigma T)} = (\Sigma T_1 - \Sigma T_0).\overline{w}_0 = (253 - 260).4,808 = -33,656 \; u.m.$$

$$I^{Q(\overline{w})} = \frac{\overline{w}_1 \Sigma T_1}{\overline{w}_0 \Sigma T_1} = 0,94531 \Rightarrow 94,53\% \Rightarrow R^{Q(\overline{w})} = -5,47\%$$

$$\Delta^{Q(\overline{w})} = (\overline{w}_1 - \overline{w}_0).\Sigma T_1 = (4,545 - 4,808).253 = -66,539 \; u.m.$$

Verificare:

$$I^{Q(\overline{w},\Sigma T)} = I^{\overline{w}} \times I^{\Sigma T} = 0,9453.0,9731 = 0,9199 \cong 0,92$$

8.b. Descompunerea producției totale pe trei factori de influență:

$$I^{Q(w,g^T,\Sigma T)} = \frac{\Sigma w_1 g_1^T \Sigma T_1}{\Sigma w_0 g_0^T \Sigma T_1} = 0,92 \Rightarrow 92\% \Rightarrow$$

$$\Rightarrow \Delta^{Q(w,g^T,\Sigma T)} \approx 100 \; u.m.$$

$$I^{Q(\Sigma T)} = \frac{\Sigma w_0 g_0^T \Sigma T_1}{\Sigma w_0 g_0^T \Sigma T_0} = \frac{\overline{w}_0 \Sigma T_1}{\overline{w}_0 \Sigma T_0} = 0,9731 \Rightarrow 97,31\%$$

$$\Delta^{Q(\Sigma T)} = \Sigma T_1 - \Sigma T_0).\overline{w}_0 33,656 \; u.m.$$

$$I^{Q(g^T)} = \frac{\Sigma w_0 g_1^T \Sigma T_1}{\Sigma w_0 g_0^T \Sigma T_1} = \frac{\overline{w}_* \Sigma T_1}{\overline{w}_0 \Sigma T_1} = 0,9886 \Rightarrow 98,86\%$$

$$\Delta^{Q(g^T)} = (\overline{w}_* - \overline{w}_0).\Sigma T_1 = -0,055.253 = -13,915 \; u.m.$$

Material cu drept de autor

$$I^{Q(w)} = \frac{\Sigma w_1 g_1^T \Sigma T_1}{\Sigma w_0 g_1^T \Sigma T_1} = \frac{\overline{w}_1 \Sigma T_1}{\overline{w}_* \Sigma T_1} = 0{,}9562 \Rightarrow 95{,}62\%$$

$$\Delta^{Q(w)} = (\overline{w}_1 - \overline{w}_*).\Sigma T_1 = -0{,}208.253 = -52{,}624\ u.m.$$

9. Valoarea producției în cele două perioade în funcție de productivitatea medie şi de numărul de salariați:

$$\Sigma p_0 q_0 = \overline{w}_0.\Sigma T_0 = 4{,}808.260 = 1250\ u.m.$$

$$\Sigma p_0 q_1 = \overline{w}_1.\Sigma T_1 = 4{,}545.253 = 1150\ u.m.$$

. Exemplu de calcul 10.2. Se cunosc datele:

Tabelul 10.13

Firma	Număr salariați		Salariul mediu, milioane lei	
	0	1	0	1
I	500	400	1,2	1,32
II	300	300	1,4	1,19
III	250	150	1,3	1,30
IV	350	400	1,25	1,10
Total	**1400**	**1250**	-	-

a. Să se analizeze dinamica şi modificarea absolută a numărului de salariați, salariul mediu şi fondului de salarizare la nivelul fiecărei firme, cu evidențierea factorilor;

b. Să se analizeze dinamica şi modificarea absolută a numărului de salariați, salariului mediu şi fondului de salarizare la nivelul ansamblului celor patru firme, cu evidențierea factorilor.

Rezolvare.

a. Pentru rezolvarea problemei se calculează fondul de salarizare pe baza relației: **FS = S.N**, coloana 2 din tabelul 10.14.

Tabelul 10.14

Fond salarizare, (milioane lei)		$i_{1/0}^{N}$	$i_{1/0}^{S}$	$i_{1/0}^{FS}$	$\Delta_{1/0}^{N}$ (persoane)
0	1				
1	**2**	**3**	**4**	**5**	**6**
600	528	0,8	1,1	0,88	-100
420	357	1,0	0,85	0,85	0
325	195	0,6	1,0	0,6	-100
437,5	440	1,143	0,88	1,01	50
Σ=1782,5	Σ= 1520	-		-	Σ = -150

Material cu drept de autor

Tabelul 10.14 (continuare)

$\Delta_{1/0}^{S}$ (mil.lei)	$\Delta_{1/0}^{FS}$ (mil.lei)	$\Delta_{1/0}^{FS(S)}$ (mil.lei)	$\Delta_{1/0}^{FS(N)}$ (mil.lei)	g_1^N	$S_0 g_1^N$ (mil.lei)
7	**8**	**9**	**10**	**11**	**12**
0,12	-72	48	-120	0,32	0,384
-0,21	-63	-63	0	0,24	0,336
0	-130	0	-130	0,12	0,156
-0,15	2,5	-60	62,5	0,32	0,400
-	Σ = -262,5	Σ = **-75**	Σ = **-187,5**	Σ = 1,00	Σ = **1,276**

Pentru firma I dinamica indicatorilor este:

$i_{1/0}^{N} = \frac{N_1}{N_0} = \frac{400}{500} = 0,8 \ (80\%)$. Numărul salariaților a scăzut în această firmă cu 20%.

$i_{1/0}^{S} = \frac{S_1}{S_0} = \frac{1,32}{1,2} = 1,1 \ (110\%)$. Salariul mediu al angajaților din firma I a crescut cu 10%.

$i_{1/0}^{FS} = \frac{FS_1}{FS_0} = \frac{528}{600} = 0,88 \ (88\%)$. Fondul de salarizare a scăzut în momentul 1 față de momentul 0 cu 12%.

Modificarea absolută a indicatorilor este:

Efectivul salariaților s-a redus cu 100 persoane:

$$\Delta_{1/0}^{N} = N_1 - N_0 = 400 - 500 = -100 \, persoane$$

Salariul mediu a crescut în momentul 1 față de momentul 0 cu 120 mii lei:

$$\Delta_{1/0}^{S} = S_1 - S_0 = 1,32 - 1,2 = 0,12 (120 \ \ mii \ lei)$$

Fondul de salarizare a scăzut, în aceeaşi perioadă, cu 72 milioane lei:

$$\Delta_{1/0}^{FS} = FS_1 - FS_0 = 528 - 600 = -72 \, mil. \, lei$$

$$\Delta_{1/0}^{FS} = FS_1 - FS_0 = S_1 N_1 - S_0 N_0 = 528 - 600 = -72 \, mil. \, lei$$

$$\Delta_{1/0}^{FS(S)} = S_1 N_1 - S_0 N_1 = \Delta_{1/0}^{S} N_1 = 0,12 \times 400 = 48 \, mil. \, lei$$

$$\Delta_{1/0}^{FS(N)} = S_0 N_1 - S_0 N_0 = S_0 \Delta_{1/0}^{N} = 12 \times (-100) = -120 \, mil. \, lei$$

Material cu drept de autor

$$\Delta_{1/0}^{FS(S)} + \Delta_{1/0}^{FS(N)} = \Delta_{1/0}^{FS} = 48 + (-120) = -72\,mil.\,lei$$

Deşi salariul mediu la firma I a crescut, la fondul de salarizare s-a înregistrat o reducere datorată scăderii numărului de angajați.

Analiza se efectuează similar pentru firmele II, III şi IV, coloanele 3-10 din tabelul 10.14.

b. La nivel de grup se pot cumula: fondul de salarizare (Σ**FS**) şi numărul de angajați (Σ**N**).

Salariul mediu se calculează:

$$\overline{S} = \frac{\Sigma FS}{\Sigma N} = \frac{\Sigma S \times N}{\Sigma N} = \Sigma S \times g^{N}$$

$$\overline{S}_1 = \frac{\Sigma FS_1}{\Sigma N_1} = \frac{1520}{1250} = 1{,}216\,mil.\,lei \quad \overline{S}_0 = \frac{\Sigma FS_0}{\Sigma N_0} = \frac{1782{,}5}{1400} = 1{,}273\,mil.\,lei$$

Atunci, dinamica indicatorilor este:

$$I_{1/0}^{\Sigma FS} = \frac{\Sigma FS_1}{\Sigma FS_0} = \frac{1520}{1782{,}5} = 0{,}853 \;\; (85{,}3\%) \qquad I_{1/0}^{\Sigma N} = \frac{\Sigma N_1}{\Sigma N_0} = \frac{1250}{1400} = 0{,}893 \;\; (89{,}3\%)$$

$$I_{1/0}^{\overline{S}} = \frac{\overline{S}_1}{\overline{S}_0} = \frac{1{,}216}{1{,}273} = 0{,}953 \;\; (95{,}3\%)$$

Dinamica salariului mediu pe ansamblul celor 4 firme se poate descompune pe seama celor doi factori de influență: salariul mediu la nivelul firmelor şi structura angajaților, coloanele 11, 12 din tabelul 10.14.

$$I_{1/0}^{\overline{S}} = \frac{\overline{S}_1}{\overline{S}_0} = \frac{\Sigma S_1 g_1^N}{\Sigma S_0 g_0^N} = \frac{1{,}216}{1{,}273} = 0{,}953 \;\; (95{,}3\%)$$

$$I_{1/0}^{\overline{S}(S)} = \frac{\Sigma S_1 g_1^N}{\Sigma S_0 g_1^N} = \frac{\overline{S}_1}{\Sigma S_0 g_1^N} = \frac{1{,}216}{1{,}273} = 0{,}953 \;\; (95{,}3\%)$$

$$I_{1/0}^{\overline{S}(g^N)} = \frac{\Sigma S_0 g_1^N}{\Sigma S_0 g_0^N} = \frac{\Sigma S_0 g_1^N}{\overline{S}_0} = \frac{1{,}276}{1{,}273} = 1{,}002 \;\; (100{,}2\%)$$

$$I_{1/0}^{\overline{S}(S)} \times I_{1/0}^{\overline{S}(g^N)} = I_{1/0}^{\overline{S}} = 0{,}953 \times 1{,}002 = 0{,}955 \;\; (95{,}5\%)$$

Dinamica fondului de salarizare se poate descompune pe seama a doi factori de influență ($\overline{S}$ şi ΣN) sau a trei factori (S, g^N şi ΣN):

$$I_{1/0}^{\Sigma FS} = \frac{\Sigma FS_1}{\Sigma FS_0} = \frac{\overline{S}_1 \Sigma N_1}{\overline{S}_0 \Sigma N_0} = \frac{\Sigma S_1 g_1^N \Sigma N_1}{\Sigma S_0 g_0^N \Sigma N_0}$$

$$I_{1/0}^{\Sigma FS(S)} = \frac{\Sigma S_1 g_1^N \Sigma N_1}{\Sigma S_0 g_1^N \Sigma N_1} = I_{1/0}^{\overline{S}(S)} = 0{,}953$$

$$I_{1/0}^{\Sigma FS(g^N)} = \frac{\Sigma S_0 g_1^N \Sigma N_1}{\Sigma S_0 g_0^N \Sigma N_1} = I_{1/0}^{\overline{S}(g^N)} = 1{,}002$$

Material cu drept de autor

$$I_{1/0}^{\Sigma FS(S)} \times I_{1/0}^{\Sigma FS(g^N)} = I_{1/0}^{\Sigma FS(\bar{S})}$$

$$I_{1/0}^{\Sigma FS(\Sigma N)} = \frac{\Sigma S_0 g_0^N \Sigma N_1}{\Sigma S_0 g_0^N \Sigma N_0} = I_{1/0}^N = 0{,}893$$

$$I_{1/0}^{\Sigma FS(S)} \times I_{1/0}^{\Sigma FS(g^N)} \times I_{1/0}^{\Sigma FS(\Sigma N)} = I_{1/0}^{\Sigma FS(\bar{S})} \times I_{1/0}^{\Sigma FS(\Sigma N)} = I_{1/0}^{\Sigma FS}$$

$$0{,}953 \times 1{,}002 \times 0{,}893 = 0{,}955 \times 0{,}893 = 0{,}853$$

Modificarea absolută a indicatorilor analizați la nivelul ansamblului, în momentul 1, față de momentul 0 este:

$$\Delta_{1/0}^{\Sigma FS} = \Sigma FS_1 - \Sigma FS_0 = 1520 - 1782{,}5 = -262{,}5 \; mil.\; lei$$

$$\Delta_{1/0}^{\Sigma N} = \Sigma N_1 - \Sigma N_0 = 1250 - 1400 = -150 \; persoane$$

$$\Delta_{1/0}^{\bar{S}} = \bar{S}_1 - \bar{S}_0 = 1{,}216 - 1{,}273 = -0{,}057 \; mil.\; lei / pers.$$

Modificarea absolută a salariului mediu pe ansamblu se poate descompune pe seama factorilor.

$$\Delta_{1/0}^{\bar{S}} = \Sigma S_1 g_1^N - \Sigma S_0 g_0^N = -0{,}057 \; mil.\; lei / pers.$$

$$\Delta_{1/0}^{\bar{S}(S)} = \Sigma S_1 g_1^N - \Sigma S_0 g_1^N = \bar{S}_1 - \Sigma S_0 g_1^N = 1{,}216 - 1{,}276 =$$

$$= -0{,}06 \; mil.\; lei / pers.$$

$$\Delta_{1/0}^{\bar{S}(g^N)} = \Sigma S_0 g_1^N - \Sigma S_0 g_0^N = \Sigma S_0 g_1^N - \bar{S}_0 = 1{,}276 - 1{,}273 =$$

$$= -0{,}003 \; mil.\; lei / pers.$$

Aşadar, salariul mediu pe salariat a scăzut cu 57 mii lei/persoană. Pe seama salariilor la nivel de firmă a scăzut cu 60 mii lei/persoană, dar a crescut cu 3 mii lei/persoană datorită deplasării structurii angajaților către firmele cu salarii mari.

$$\Delta_{1/0}^{\bar{S}(S)} + \Delta_{1/0}^{\bar{S}(g^N)} = \Delta_{1/0}^{\bar{S}}$$

$$-0{,}06 + 0{,}003 = -0{,}057 mil.lei / pers.$$

Pentru fondul de salarizare analiza modificării absolute pe seama a 2 sau 3 factori de influență se poate face astfel:

$$\Delta_{1/0}^{\Sigma FS} = \Sigma FS_1 - \Sigma FS_0 = \bar{S}_1 \Sigma N_1 - \bar{S}_0 \Sigma N_0 =$$

$$= \Sigma S_1 g_1^N \Sigma N_1 - \Sigma S_0 g_0^N \Sigma N_0 = -262{,}5 \; mil.\; lei$$

$$\Delta_{1/0}^{\Sigma FS(S)} = \Sigma S_1 g_1^N \Sigma N_1 - \Sigma S_0 g_1^N \Sigma N_1 = \Delta_{1/0}^{\bar{S}(S)} \times \Sigma N_1 =$$

$$= -0{,}06 \times 1250 = -75 \; mil.\; lei \; (total\; col.9,\; tabelul\; 10.14$$

Material cu drept de autor

$$\Delta_{1/0}^{\Sigma FS(g^N)} = \Sigma S_0 g_1^N \Sigma N_1 - \Sigma S_0 g_0^N \Sigma N_1 =$$

$$= \Delta_{1/0}^{\bar{S}(g_1^N)} \times \Sigma N_1 = -0{,}003 \times 1250 = 3{,}5 \; mil.\; lei$$

$$\Delta_{1/0}^{\Sigma FS(\Sigma N)} = \Sigma S_0 g_0^N \Sigma N_1 - \Sigma S_0 g_0^N \Sigma N_0 =$$

$$= \bar{S}_0 \Delta_{1/0}^{\Sigma N} = 1{,}273 \times (-150) = -191 \; mil.\; lei$$

$$\Delta_{1/0}^{\Sigma FS(S)} + \Delta_{1/0}^{\Sigma FS(g_1^N)} = \Delta_{1/0}^{\Sigma FS(\bar{S})} =$$

$$= -75 + 3{,}5 = -71{,}5 \; mil.\; lei$$

$$\Delta_{1/0}^{\Sigma FS(g^N)} + \Delta_{1/0}^{\Sigma FS(\Sigma N)} = \Delta_{1/0}^{\Sigma FS(N)} =$$

$$= 3{,}5 - 191 = -187{,}5 \; mil.\; lei \; (total\; col.9, tabelul\; 10.14$$

$$\Delta_{1/0}^{\Sigma FS(S)} + \Delta_{1/0}^{\Sigma FS(g_1^N)} + \Delta_{1/0}^{\Sigma FS(\Sigma N)} =$$

$$= -75 \quad + \quad 3{,}5 \quad + \quad (.191) \; = -262{,}5 \; mil.\; lei$$

Material cu drept de autor

BIBLIOGRAFIE

1. Anghelache Constantin, *Statistică. Teorie şi aplicații*, Bucureşti, Editura Economică, 1998.
2. Baron Tudor ş. a., *Statistică teoretică şi economică*, Bucureşti, Editura Didactică şi Pedagogică, 1996.
3. Bădiță Maria, Baron Tudor, Korka Mihai, *Statistică pentru afaceri*, Editura Eficient, Bucureşti, 1998.
4. Begu Liviu-Stelian, Erika Tusa, Statistică teoretică şi economică, Bucureşti, ASE, 2004.
5. Biji E., P. Wagner, E. Lllea, N. Petcu, Statistica, Bucureşti, Editura Universitatea "Titu Maiorescu", 1995.
6. Cenuşă Gheorghe, Teoria probabilităților, Bucureşti, ASE, 2004.
7. Coman Gheorghe, Statistica (aplicații). Iaşi, PIM, 2005.
8. Coman Gh., Murgu Al., *Statistică teoretică şi aplicată*, Târgu Mureş, Editura „Dimitrie Cantemir", 2000.
9. Gérald Baillargeon, Louise Martin, Outils statistiques pour les sciences du comportement et de la psychologie (1Cédérom), Editeur: SMG, 2003.
10. Gérald Baillargeon, Louise Martin, Outils statistiques pour les sciences du comportement et de la psychologie: Corrigé des exercices, Editeur: SMG, 2003.
11. Gérald Baillargeon, Statistique appliquée pour les sciences de la gestion et les sciences économiques (1Cédérom), Paris, Editeur: SMG, 2003.
12. Gérald Baillargeon, Statistique appliquée pour les sciences de la gestion et les sciences économiques: Corrigé des exercices, Paris, Editeur: SMG, 2003.
13. Gérard Forzy, Manuel de Statistique, Editeur: Ellipses Marketing, 2005.
14. Grais B. Méthodes Statistiques , Paris, Dunod, 1992.
15. Guyon Xavier, Statistique et économétrie - Du modèle linéaire aux modèles non-linéaires, Editeur: Ellipses Marketing, Paris, 2001.
16. Haber A., Runyon R., *General Statistics*, Addison-Weslwy, Reading, Mass., 1977.
17. Hubert Egon, Pascal Porée, Statistique et probabilités en production industrielle: Volume 2, Contrôle et maîtrise de la qualité, fiabilité, problèmes et exercices corrigés, Paris, Editeur: Hermann, 2004.
18. Hubert Egon, Pascal Porée, Statistique et probabilités en production industrielle: Volume 1, Etude générale, problèmes et exercices corrigés, Paris, Editeur: Hermann, 2004.
19. Iosifescu M., Mihoc Gh., Theodorescu R., *Teoria probabilităților şi statistică matematică*, Bucureşti, Ed. Tehnică, 1966.
20. Isaic-Maniu, Eugen Pecican, Daniela Ştefănescu, Viorel Gh. Vodă, Pavel Wagner, Dicționar de statistică generală, Bucureşti, Editura Economică, 2003.
21. Isaic-Maniu Al. A. Gădinaru, V. Voineagu, C. Mitruț, *Statistică teoretică şi economică*, Chişinău, Editura Tahnică, 1994.

Material cu drept de autor

22. Ivănescu Ion ş.a., *Statistică*, Bucureşti, Editura Didactică şi Pedagogică, 1980.
23. Jaba Elisabeta, *Statistica,* Bucureşti, Editura Economică, 1998.
24. Jean-Pierre Lecoutre, Statistique et probabilités, Paris, Editeur: Dunod, 2005.
25. Jolion Jean-Michel, Probabilités et Statistique, Version électronique: **http://rfv.insa-lyon.fr/~jolion/STAT/poly.html.**
26. Kmenta J., *Elements of Econometrics*, Macmillan Publishing Co., Inc., New York, 1971.
27. Lecoutre Jean-Pierre, Statistique et probabilités, Editeur: Dunod, Paris, 2005.
28. Lilea Eugenia, Mihaela Vătui, Doina Boldeanu, Zizi Goschin, Statistica, Bucureşti, ASE, 2004.
29. Luc Albarello, Jean-Luc Guyot, Etienne Bourgeois, Statistique descriptive, Paris, Editeur: De Boeck, 2002.
30. Lucile Chanquoy, Statistiques appliquées à la psychologie et aux sciences humaines et sociales, Paris, Editeur: Hachette, 2005.
31. Maurice Lethielleux, Statistique descriptive, Paris, Editeur: Dunod, 2005.
32. Mihoc Gh., *Elemente de calculul probabilităţilor*, Bucureşti, Editura Tehnică, 1954.
33. Mihoc Gh., Micu N., *Teoria probabilităţilor şi statistică matematică*, Bucureşti, Editura Didactică şi Pedagogică, 1980.
34. Negoescu Gheorghe, Rodica Ciobanu, Cristina-Aurora Bontaş, *Bazele statisticii pentru afaceri*, Bucureşti, Editura ALL BECK, 1999.
35. Onicescu O., *Curs de teoria probabilităţilor*, Bucureşti, Ed. Tehnică, 1956.
36. Păun Mihai, Carmen Hartulari, Analiza, diagnoza si evaluarea sistemelor din economie, Bucureşti, ASE, 2003.
37. Philippe Michel, Cours de mathématiques pour économistes, Paris, Editeur: Economica, 1989.
38. Philippe Tassi, Méthodes statistiques, Paris, Editeur: Economica, 2004.
39. Rancu N., Tovissi L., *Statistică matematică cu aplicaţii în producţie*, Bucureşti, Editura Academiei RPR, 1963.
40. Roman Monica, Statistica financiar-bancară, Bucureşti, ASE, 2005.
41. Saporta G. Théorie et méthodes de la statistique, Paris, Technip, 1987.
42. Soliani Lamberto, MANUALE DI STATISTICA PER LA RICERCA E LA PROFESSIONE (edizione aprile 2005),Versiunea electronică: **http://www.dsa.unipr.it/soliani/soliani.html.**
43. Şerban Daniela, Statistica pentru studii de marketing si administrarea afacerilor, Bucureşti, ASE, 2004.
44. Tassi Philippe, Méthodes statistiques, Editeur: Economica, Paris, 2004.
45. Ţiţian Emilia, Simona Ghiţă, Cristina Trandas, Statistica economică, Bucureşti, ASE, 2004.
46. Troie Liviu, Octavian Zaharia, Monica Roman, Miruna Hurduzeu, Analiza statistica a activităţii economice şi a gestiunii financiare a întreprinderii, Bucureşti, ASE, 2002.

Material cu drept de autor

47. Vasilescu Gh., Niculescu I., Wagner Fl., Zaharia O., Analiza statistico-economică în industrie, Bucureşti, Editura Didactică şi Pedagogică, 1997.
48. Vasilescu Gh., Wagner Fl., Zaharia O., Roman M., Hurduzeu M., Analiza statistico-economică în industrie. Culegere de probleme şi teste grilă, Bucureşti, ASE, 1999.
49. Vodă V. Gheorghe, *Gândirea statistică, un mod de gândire al viitorului*, Bucureşti, Editura Albatros, 1977.
50. Voineagu Mariana, Emilia Țițian, Simona Ghiță, *Statistică aplicată*, Bucureşti, Editura Fundației „România de Mâine", 2000.
51. Voineagu Virgil, Constantin Mitruț, Emilia Țițian, Simona Ghiță, Statistica, Bucureşti, ASE, 2004.
52. Zaharia Octavian, Aniela Danciu, Monica Roman, Statistica întreprinderii, Bucureşti, ASE, 2004.

Material cu drept de autor

Anexa 1

Valorile funcției lui Gauss

$$f(z)=\frac{1}{\sqrt{2.\pi}}e^{-\frac{z^2}{2}};\qquad 0{,}00<z<3{,}29\qquad f(-z)=f(z)$$

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3957	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3726	0,3712	0,3697
0,4	0,3683	0,3668	0,3653	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2613	0,2603	0,2589	0,2565	0,2541	0,2516	0,2429	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0396	0,0388	0,0379	0,0371	0,0363
2,2	0,0355	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
2,9	0,060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018

Material cu drept de autor

Anexa 2

Funcția lui Laplace

$$\Phi(z) = \frac{1}{\sqrt{2.\pi}} \int_0^z e^{-\frac{t^2}{2}} .dt \quad 0,00 < z < 3,09 \qquad \Phi(-z) = \Phi(z)$$

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0909	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1555	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2045	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2557	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2703	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3683	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4865	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4078	0,4879	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4986	0,4987	0,4987	0,4988	0,4988	0,4990	0,4991	0,4991	0,4992	0,4993

Material cu drept de autor

Anexa 3

Repartiția Student

Valorile lui t în funcție de probabilitățile **$P(t<+t_p) = 1-q = P$** şi numărul gradelor de libertate ν

ν \ 1-q	0,995	0,99	0,975	0,95	0,90	0,85	0,80
1	63,657	31,821	12,706	6,314	3,078	1,963	1,376
2	9,925	6,965	4,303	2,920	1,886	1,386	1,061
3	5,841	4,541	3,182	2,353	1,638	1,250	0,978
4	4,604	3,747	2,776	2,132	1,533	1,190	0,941
5	4,032	3,365	2,571	2,015	1,476	1,156	0,920
6	3,707	3,143	2,447	1,943	1,440	1,134	0,906
7	3,499	2,998	2,365	1,895	1,415	1,119	0,896
8	3,355	2,896	2,306	1,860	1,397	1,108	0,889
9	3,250	2,821	2,262	1,833	1,383	1,100	0,883
10	3,169	2,764	2,228	1,812	1,372	1,093	0,879
11	3,106	2,718	2,201	1,796	1,363	1,088	0,876
12	3,055	2,681	2,179	1,782	1,356	1,083	0,873
13	3,012	2,650	2,160	1,771	1,350	1,079	0,870
14	2,977	2,624	2,145	1,761	1,345	1,076	0,868
15	2,947	2,602	2,131	1,753	1,341	1,074	0,866
16	2,921	2,583	2,120	1,746	1,337	1,071	0,865
17	2,898	2,567	2,110	1,740	1,333	1,069	0,863
18	2,878	2,552	2,101	1,734	1,330	1,067	0,862
19	2,861	2,539	2,093	1,729	1,328	1,06	0,861
20	2,845	2,528	2,085	1,727	1,325	1,064	0,860
21	2,831	2,518	2,080	1,721	1,323	1,063	0,859
22	2,819	2,508	2,074	1,717	1,321	1,061	0,858
23	2,807	2,500	2,069	1,714	1,319	1,060	0,858
24	2,797	2,492	2,064	1,711	1,318	1,059	0,857
25	2,787	2,485	2,060	1,708	1,316	1,058	0,856
26	2,779	2,479	2,056	1,706	1,315	1,058	0,856
27	2,771	2,473	2,052	1,703	1,314	1,057	0,855
28	2,763	2,467	2,048	1,701	1,313	1,056	0,855
29	2,756	2,462	2,045	1,699	1,311	1,055	0,854
30	2,750	2,457	2,042	1,697	1,310	1,055	0,854
40	2,704	2,425	2,021	1,684	1,303	1,050	0,851
60	2,660	2,390	2,000	1,671	1,296	1,046	0,848
120	2,617	2,358	1,980	1,658	1,289	1,041	0,845
∞	2,576	2,326	1,960	1,645	1,282	1,036	0,842

Material cu drept de autor

Anexa 4

Valorile r_{ij} pentru testul lui Dixon

Nr. selecției	r_{10}	r_{11}	r_{12}	r_{20}	r_{21}	r_{22}
4	0,926	0,995	-	0,996	-	-
5	0,821	0,937	0,996	0,950	0,998	-
6	0,740	0,839	0,951	0,865	0,970	0,998
7	0,680	0,782	0,875	0,814	0,919	0,970
8	0,634	0,725	0,797	0,746	0,868	0,922
9	0,598	0,677	0,739	0,700	0,816	0,873
10	0,568	0,639	0,694	0,664	0,760	0,826
11	0,542	0,606	0,658	0,627	0,713	0,781
12	0,522	0,580	0,629	0,612	0,675	0,740
13	0,503	0,558	0,612	0,590	0,649	0,705
14	0,488	0,539	0,580	0,571	0,627	0,674
15	0,475	0,522	0,560	0,554	0,607	0,647
16	0,463	0,508	0,544	0,539	0,589	0,624
17	0,452	0,495	0,529	0,526	0,573	0,605
18	0,442	0,484	0,516	0,514	0,559	0,589
19	0,433	0,473	0,504	0,503	0,547	0,575
20	0,425	0,464	0,493	0,494	0,536	0,562
21	0,18	0,455	0,483	0,485	0,526	0,551
22	0,411	0,446	0,474	0,477	0,517	0,541
23	0,404	0,439	0,465	0,469	0,509	0,532
24	0,399	0,432	0,457	0,462	0,501	0,524
25	0,393	0,426	0,450	0,456	0,493	0,516

Material cu drept de autor

Anexa 5

Valorile raportului $F(\nu_1;\nu_2)=\dfrac{S_1^2}{S_2^2}$ corespunzător probabilității **P(F ≤ F_p)=0,95**

și numerelor gradelor de libertate ν_1 și ν_2, $S_1^2 > S_2^2$

(extras)

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10
1	161	200	216	225	230	234	237	239	241	242
2	18,5	19,0	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,37	3,29	3,23	3,18	3,14	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98

Material cu drept de autor

Anexa 6

Testul de semnificație după Cochran

v / k	1	2	3	4	5	6	7	8	9	10
q = 0,05										
2	0,999	0,975	0,939	0,906	0,877	0,853	0,833	0,816	0,801	0,788
3	0,967	0,871	0,798	0,746	0,707	0,677	0,653	0,633	0,617	0,603
4	0,907	0,768	0,684	0,629	0,590	0,560	0,537	0,519	0,507	0,488
5	0,841	0,684	0,598	0,544	0,507	0,478	0,456	0,439	0,424	0,412
6	0,781	0,616	0,532	0,480	0,445	0,418	0,348	0,382	0,368	0,357
7	0,727	0,561	0,480	0,431	0,397	0,373	0,354	0,338	0,326	0,315
8	0,680	0,516	0,438	0,391	0,360	0,336	0,319	0,304	0,293	0,283
9	0,639	0,478	0,403	0,358	0,329	0,307	0,290	0,277	0,266	0,257
10	0,602	0,445	0,373	0,331	0,303	0,282	0,267	0,254	0,244	0,235
q = 0,01										
2	0,999	0,995	0,979	0,959	0,937	0,917	0,900	0,882	0,867	0,854
3	0,993	0,942	0,883	0,834	0,793	0,761	0,734	0,711	0,691	0,674
4	0,968	0,864	0,781	0,721	0,676	0,641	0,613	0,590	0,570	0,554
5	0,928	0,789	0,696	0,633	0,588	0,553	0,526	0,504	0,485	0,470
6	0,883	0,722	0,626	0,564	0,520	0,487	0,461	0,440	0,423	0,408
7	0,834	0,664	0,569	0,508	0,466	0,485	0,411	0,391	0,375	0,361
8	0,795	0,615	0,521	0,462	0,423	0,370	0,352	0,370	0,352	0,325
9	0,754	0,573	0,481	0,425	0,387	0,359	0,338	0,321	0,307	0,295
10	0,718	0,536	0,447	0,393	0,357	0,331	0,311	0,295	0,281	0,270

Material cu drept de autor

Material cu drept de autor

Anexa 6

Valorile probabilităților $S_n(t)$ pentru distribuția Student

t	n_1														
	2	3	4	5	6	7	8	9	10	12	14	16	18	20	α
	$S_n(t)$														
0,0	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,50000
0,2	0,563	0,570	0,573	0,574	0,575	0,576	0,576	0,577	0,577	0,577	0,578	0,578	0,578	0,578	0,57926
0,4	0,621	0,636	0,642	0,645	0,647	0,648	0,650	0,650	0,651	0,652	0,652	0,653	0,653	0,653	0,65542
0,6	0,672	0,695	0,705	0,710	0,713	0,715	0,716	0,717	0,718	0,720	0,721	0,721	0,722	0,722	0,72575
0,8	0,715	0,746	0,759	0,766	0,770	0,773	0,775	0,777	0,778	0,780	0,781	0,782	0,783	0,783	0,78814
1,0	0,750	0,788	0,804	0,813	0,818	0,822	0,825	0,827	0,828	0,831	0,832	0,833	0,834	0,835	0,84134
1,2	0,779	0,824	0,842	0,852	0,858	0,862	0,865	0,868	0,870	0,872	0,874	0,876	0,877	0,878	0,88493
1,4	0,803	0,852	0,872	0,883	0,890	0,894	0,898	0,900	0,902	0,906	0,908	0,909	0,910	0,911	0,91924
1,6	0,822	0,875	0,896	0,908	0,915	0,920	0,923	0,926	0,928	0,931	0,933	0,935	0,936	0,937	0,94520
1,8	0,839	0,893	0,915	0,927	0,934	0,939	0,943	0,945	0,947	0,950	0,952	0,954	0,955	0,956	0,96407
2,0	0,852	0,908	0,930	0,942	0,949	0,954	0,957	0,960	0,962	0,965	0,967	0,968	0,969	0,970	0,97725
2,2	0,864	0,921	0,942	0,954	0,960	0,965	0,968	0,970	0,972	0,975	0,977	0,978	0,979	0,980	0,98610
2,4	0,874	0,931	0,952	0,963	0,969	0,973	0,976	0,978	0,980	0,982	0,984	0,985	0,986	0,987	0,99180
2,6	0,883	0,938	0,960	0,970	0,976	0,980	0,982	0,984	0,986	0,988	0,989	0,990	0,991	0,991	0,99534
2,8	0,891	0,946	0,966	0,976	0,981	0,984	0,987	0,988	0,990	0,991	0,992	0,993	0,994	0,994	0,99744
3,0	0,898	0,952	0,971	0,980	0,985	0,988	0,990	0,992	0,992	0,994	0,995	0,996	0,996	0,966	0,99865
3,2	0,904	0,957	0,957	0,984	0,988	0,991	0,992	0,994	0,995	0,996	0,996	0,997	0,997	0,998	0,99931
3,4	0,909	0,962	0,979	0,986	0,990	0,993	0,994	0,995	0,996	0,997	0,998	0,998	0,998	0,998	0,99966
3,6	0,914	0,965	0,982	0,989	0,992	0,994	0,996	0,996	0,997	0,998	0,998	0,999	0,999	0,999	0,99984
38	0,918	0,969	0,986	0,990	0,994	0,996	0,997	0,997	0,998	0,998	0,999	0,999	0,999	0,999	0,99993
4,0	0,922	0,971	0,986	0,992	0,995	0,996	0,997	0,998	0,998	0,999	0,999	0,999	1,000	1,000	0,99997
4,2	0,926	0,974	0,988	0,993	0,996	0,997	0,998	0,998	0,999	0,999	1,000	1,000	-	-	0,99999
4,4	0,929	0,976	0,989	0,994	0,996	0,998	0,998	0,999	0,999	1,000	-	-	-	-	0,99999
4,6	0,932	0,978	0,990	0,995	0,997	0,998	0,999	0,999	0,999	-	-	-	-	-	-
4,8	0,935	0,980	0,991	0,996	0,998	0,998	0,999	0,999	1,000	-	-	-	-	-	-
5,0	0,937	0,981	0,992	0,996	0,998	0,999	0,999	1,000	-	-	-	-	-	-	-

Material cu drept de autor

Anexa 7

Valorile funcției **L (q,k)**

k	q = ε/S																	
	0,10	0,15	0,20	0,25	0,30	0,40	0,50	0,60	0,70	0,80	0,90	1,00	1,25	1,50	1,75	2,00	2,50	3,00
	L (q,k)																	
6	0,264	0,388	0,501	0,599	0,681	0,791	0,849	0,886	0,913	0,933	0,948	0,959	0,978	0,987	0,992	0,995	0,998	0,999
8	0,305	0,444	0,567	0,669	0,748	0,845	0,895	0,926	0,948	0,963	0,974	0,981	0,991	0,996	0,998	0,999	1,000	1,000
10	0,340	0,491	0,620	0,722	0,797	0,882	0,925	0,961	0,968	0,979	0,986	0,991	0,997	0,999	0,999	1,000	-	-
12	0,371	0,532	0,664	0,764	0,833	0,900	0,946	0,968	0,980	0,988	0,993	0,996	0,999	1,000	1,000	-	-	-
14	0,399	0,567	0,701	0,798	0,862	0,929	0,960	0,978	0,988	0,993	0,996	0,998	0,999	-	-	-	-	-
16	0,425	0,599	0,733	0,826	0,885	0,944	0,971	0,985	0,992	0,996	0,998	0,999	1,000	-	-	-	-	-
18	0,448	0,627	0,760	0,849	0,903	0,955	0,980	0,990	0,996	0,998	0,999	0,999	-	-	-	-	-	-
20	0,470	0,652	0,784	0,868	0,918	0,964	0,984	0,993	0,997	0,999	0,999	1,000	-	-	-	-	-	-
25	0,518	0,706	0,832	0,905	0,944	0,979	0,992	0,997	0,999	1,000	1,000	-	-	-	-	-	-	-
30	0,559	0,749	0,867	0,930	0,962	0,988	0,996	0,999	1,000	-	-	-	-	-	-	-	-	-
35	0,597	0,787	0,893	0,944	0,969	0,990	0,997	0,999	-	-	-	-	-	-	-	-	-	-
40	0,628	0,815	0,913	0,957	0,978	0,994	0,999	1,000	-	-	-	-	-	-	-	-	-	-
45	0,657	0,840	0,929	0,967	0,984	0,996	0,999	-	-	-	-	-	-	-	-	-	-	-
50	0,682	0,860	0,942	0,974	0,993	0,998	0,999	-	-	-	-	-	-	-	-	-	-	-
60	0,726	0,893	0,960	0,984	0,996	0,999	1,000	-	-	-	-	-	-	-	-	-	-	-
70	0,762	0,917	0,972	0,990	0,998	1,000	-	-	-	-	-	-	-	-	-	-	-	-
80	0,972	0,935	0,980	0,994	0,999	-	-	-	-	-	-	-	-	-	-	-	-	-
90	0,818	0,949	0,986	0,996	0,999	-	-	-	-	-	-	-	-	-	-	-	-	-
100	0,840	0,959	0,990	0,997	1,000	-	-	-	-	-	-	-	-	-	-	-	-	-
150	0,914	0,986	0,988	1,000	-	-	-	-	-	-	-	-	-	-	-	-	-	-
200	0,951	0,995	1,000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
250	0,972	0,998	1,000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
500	0,998	1,000	1,000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1000	1,000	1,000	1,000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Material cu drept de autor

Anexa 8

Distribuția χ^2. Valorile lui χ^2 în funcție de probabilitățile **P =** $P\left(\chi^2 \leq \chi_P^2\right)$ și numărul gradelor de libertate

P / ν	Probabilitatea													
	0,99	0,98	0,95	0,90	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
2	0,0201	0,0404	0,103	0,211	0,446	0,713	1,386	2,408	3,219	4,605	5,991	7,384	9,210	13,815
3	0,115	0,185	0,352	0,584	1,005	1,424	2,366	3,665	4,642	6,251	7,815	9,837	11,345	16,266
4	0,297	0,429	0,711	1,064	1,649	2,195	3,357	4,878	5,989	7,779	9,488	11,668	13,227	18,467
5	0,554	0,752	1,145	1,610	2,343	3,000	4,351	6,064	7,289	9,236	11,070	13,388	15,086	20,515
6	0,872	1,134	1,635	2,204	3,070	3,828	5,348	7,231	8,558	10,645	12,592	15,033	16,812	22,457
7	1,239	1,564	2,167	2,833	3,822	4,671	6,346	8,383	9,803	12,017	14,067	16,622	18,475	24,322
8	1,646	2,032	2,733	3,490	4,594	5,527	7,344	9,524	11,030	13,462	15,507	18,168	20,090	26,125
9	2,088	2,532	3,325	4,168	5,380	6,393	8,343	10,656	12,242	14,684	16,919	19,679	21,666	27,877
10	2,558	3,059	3,940	4,865	6,179	7,267	9,342	11,781	13,442	15,987	18,307	21,161	23,209	29,588
11	3,053	3,609	4,575	5,578	6,989	8,148	10,341	12,899	14,631	17,275	19,675	22,618	24,725	31,264
12	3,571	4,178	5,226	6,304	7,807	9,034	11,340	14,011	15,812	18,549	21,026	24,054	26,217	32,909
13	4,107	4,765	5,892	7,042	8,634	9,926	12,340	15,119	16,985	19,812	22,362	25,472	27,688	34,528
14	4,660	5,368	6,571	7,790	9,467	10,821	13,339	16,222	18,151	21,064	23,685	26,873	29,141	36,123
15	5,229	5,985	7,261	8,547	10,307	11,721	14,939	17,322	19,311	22,307	24,996	28,259	30,578	37,697
16	5,812	6,614	7,962	9,312	11,152	12,624	15,338	18,418	20,465	23,542	26,296	29,633	32,000	39,252
17	6,408	7,255	8,672	10,085	12,002	13,351	16,338	19,511	21,615	24,769	27,587	30,995	33,409	40,790
18	7,015	7,906	9,390	10,865	12,857	14,440	17,338	20,601	22,760	25,989	28,869	32,346	34,805	42,312
19	7,633	8,567	19,117	11,651	13,716	15,352	18,338	21,689	23,900	27,204	30,144	33,687	36,191	43,820

Material cu drept de autor

Anexa 8 (continuare)

20	8,260	9,237	10,851	12,443	14,578	16,266	19,337	22,775	25,038	28,412	31,410	35,020	37,566	45,315
21	8,897	9,915	11,591	13,240	15,445	17,182	20,337	23,858	26,171	29,615	32,671	36,343	38,932	46,797
22	9,542	10,600	12,338	14,041	16,314	18,101	21,337	24,939	27,301	30,813	33,924	37,659	40,289	48,268
23	10,196	11,293	13,091	14,848	17,187	19,021	22,337	26,018	28,429	32,007	35,172	38,968	41,638	49,628
24	10,856	11,992	13,848	15,659	18,062	19,943	23,337	27,096	29,553	33,196	36,415	40,270	42,980	51,179
25	11,524	12,697	14,611	16,473	18,940	20,867	24,337	28,172	30,675	34,652	37,652	41,566	44,314	52,620

Material cu drept de autor

Gheorghe COMAN
CURRICULUM VITAE

M-am născut la 20 martie 1933, în Comuna Scorțaru Nou, Județul Brăila, într-o familie de țărani.

În 1960 am absolvit Facultatea de Mecanică din Institutul Politehnic "Gheorghe Asachi" Iaşi şi datorită situației şcolare foarte bune am fost încadrat în învățământ la Catedra de Tehnologia Metalelor din Facultatea de Mecanică, Institutul Politehnic "Gheorghe Asachi" Iaşi.

În perioada 1 octombrie 1961 – 1 octombrie 1964 am fost încadrat asistent la Catedra de Tehnologia Metalelor cu sarcini didactice la disciplinele: ***Tehnologie mecanică; Tehnologia materialelor; Studiul metalelor; Tehnologia construcției de maşini; Tehnologia fabricației maşinilor termice; Bazele tehnologiei construcției de maşini; Tehnologia matrițării şi ştanțării la rece; Atelier mecanic***.

În perioada 1 octombrie 1964 – 1 octombrie 1969 am fost încadrat asistent cu delegație de predare la Catedra de Tehnologia construcției de maşini şi Mecanică agricolă, cu sarcini didactice la disciplinele: ***Tehnologia construcției maşinilor-unelte; tehnologia construcției de maşini; Tehnologia matrițării şi ştanțării la rece***.

În perioada 1 octombrie 1969 – 9 februarie 1977 am fost încadrat şef de lucrări, prin concurs, la Catedra de Tehnologia construcției de maşini şi Mecanică agricolă, cu sarcini didactice la disciplinele: ***Tehnologia construcției maşinilor-unelte; Tehnologia construcției de maşini; Tehnologia reparării utilajului agricol; Procese tehnologice speciale; Tehnologia fabricării maşinilor***.

La 9 februarie 1977 am fost încadrat conferențiar, prin concurs, la Catedra de Tehnologia construcției de maşini şi mecanică agricolă cu sarcini didactice la disciplinele: ***Tehnologia construcției maşinilor-unelte; Tehnologii neconvenționale; Bazele cercetării experimentale***.

La 15 septembrie 1978 am fost încadrat conferențiar şef de catedră la Catedra de Tehnologia metalelor cu sarcini didactice la disciplinele: ***Tehnologia construcției maşinilor-unelte; Tehnologia materialelor; Studiul metalelor; Metalurgia pulberilor; Tehnologia fabricării şi reparării utilajului tehnologic***.

În 1982 am fost ales şef de catedră, iar în 1986 şi ianuarie 1990 am fost reales şef de catedră la Catedra de Tehnologia Metalelor, fiind în această funcție până la 1 octombrie 1990.

La disciplinele menționate am ținut prelegeri, am condus proiecte de an şi diplomă, am efectuat lucrări practice şi am condus cercuri ştiințifice studențeşti.

Material cu drept de autor

Între 1977 şi 1980 am fost, în fiecare sesiune, membru în Comisia de Examen de Stat, iar între 1980 şi 1990 am fost, în fiecare sesiune, preşedinte de Comisie de Examen de Stat.

La disciplinele la care am avut sarcini didactice m-am preocupat permanent de îmbunătățirea continuă a prelegerilor prin introducerea noutăților ştiințifice, fiind permanent la curent cu noile descoperiri ştiințifice în domeniile respective pe plan mondial, introducerea unor lucrări de laborator cu un conținut ştiințifico-didactic cât mai complex, îmbunătățirea continuă a conținutului proiectelor de an şi diplomă, pe baza rezolvării unor teme ce interesau practica productivă din întreprinderile constructoare de maşini din țara noastră, precum şi prin efectuarea unor lucrări cu caracter teoretico-experimental în cadrul cercurilor ştiințifice studențeşti, preocupări puse în evidență de conținutul manualelor şi îndrumarelor elaborate pentru studenți, inclusiv cel de față.

Între 1990 - 1992 am colaborat la Universitatea Ecologică "Dimitrie Cantemir" în calitate de profesor asociat la disciplinele: ***Economia cercetării şi modernizării produselor industriale; Analiza valorii*** şi ***Statistica***.

Între 1992-1995 am colaborat la organizarea Universității "George Bacovia" Bacău fiind profesor asociat la disciplinele: ***Ecologie globală (Economia mediului), Analiza valorii*** şi ***Statistica***. Am îndeplinit şi funcția de Rector la autorizarea ei.

Între 1995-2005 am fost profesor asociat la Universitatea Ecologică "Dimitrie Cantemir" Iaşi la disciplinele: ***Economia mediului; Analiza valorii; Ecologie spirituală*** şi ***Statistica***.

Din anul 2003 sunt profesor asociat la Universitatea „Ştefan Lupaşcu" Iaşi, la disciplinele: ***Ecologie spirituală; Economia mediului; Statistica; Econometrie***.

La toate aceste discipline am manuale elaborate. Aceasta cred că este o obligație morală a oricărui cadru didactic, de a pune la dispoziția studenților propriul manual, conform cerinței elementare că nu este moral a fi exigent cu alții dacă nu eşti exigent cu tine însuți.

La 15 martie 1975 am susținut teza de doctorat cu tema ***"Contribuții privind transferul erorii de bazare pe suprafața prelucrată la rectificarea fără centre cu bazarea semifabricatelor pe reazeme fixe"***, conducător ştiințific prof. dr. ing. Constantin Picoş.

Până în prezent activitatea mea ştiințifică este concretizată în următoarele realizări:

- peste 60 de cărți publicate: manuale, îndrumare, tratate, monografii (31 în edituri de interes național şi peste 35 de interes local, destinate activității didactice cu studenții);
- 52 de articole în reviste de specialitate din țară şi străinătate;
- 84 de lucrări comunicate la diferite sesiuni ştiințifice tematice şi publicate în volume editate cu aceste ocazii;
- 3 recenzii;
- 4 descrieri de invenții.

Material cu drept de autor

Începând cu anul 1969, toată activitatea mea ştiinţifică s-a desfăşurat pe bază de contracte de cercetare încheiate cu diferite întreprinderi constructoare de maşini din ţară. Am fost titular la circa 30 de contracte de cercetare ştiinţifică, cu o valoare de peste 10 milioane de lei (preţuri înainte de 1989).

Menţionez următoarele cărţi publicate în edituri de nivel naţional:

1. ***Probleme actuale ale finisării şi suprafinisării suprafeţelor pieselor de maşini. Finisarea pieselor de maşini***, Bucureşti, INID, 1973, vol.1, 124 p.
2. ***Probleme actuale ale finisării şi suprafinisării suprafeţelor pieselor de maşini. Suprafinisarea suprafeţelor pieselor de maşini***, Bucureşti, INID, 1973, vol.2, 104 p.
3. ***Calculul adausurilor de prelucrare şi al regimurilor de aşchiere***, Bucureşti, Editura Tehnică, 1974, 603 p.
4. ***Tehnologia construcţiei de maşini. Probleme***, Bucureşti, Editura Didactică şi Pedagogică, 1976, 400 p.
5. ***Normarea tehnică pentru prelucrări prin aşchiere***, Bucureşti, Editura Tehnică, vol.1, 1979, 336 p.
6. ***Prelucrabilitatea prin aşchiere a aliajelor feroase***, Bucureşti, Editura Tehnică, 1981, 242 p.
7. ***Normarea tehnică pentru prelucrări prin aşchiere***, Bucureşti, Editura Tehnică, vol.2, 1982, 208 p.
8. ***Rulmenţi. Proiectare şi tehnologie***, Bucureşti, Editura tehnică, 1985, 391 p.
9. ***Proiectarea tehnologiilor de prelucrare mecanică prin aşchiere. Manual de proiectare***, Vol.1, Chişinău, Editura Universitas, 1992, 640 p., ISBN 5-362-00970-2.
10. ***Proiectarea tehnologiilor de prelucrare mecanică prin aşchiere. Manual de proiectare***, Vol.2, Chişinău, Editura Universitas, 1992, 408 p., ISBN 5-362-00971-0.
11. ***Economia mediului***, Iaşi, Editura Moldovița, 1996, 348 p., ISBN 973-95206-2-8.
12. ***Tehnologia proceselor productive***, Iaşi, Editura Moldovița, 1996, 200 p., ISBN 973-95206-3-8.
13. ***Tehnologia fabricaţiei produselor industriale***, Târgu Mureş, Editura "Dimitrie Cantemir", 1999, 214 p., ISBN 973-8042-03-8.
14. ***Analiza valorii***, Târgu Mureş, Editura "Dimitrie Cantemir", 2000, 340 p., ISBN 973 – 8042 – 09 – 7.
15. ***Economia mediului***, Târgu Mureş, Editura "Dimitrie Cantemir", 2000, 290 p., ISBN 973- 99596 – 6 – 0.
16. ***Statistică teoretică şi aplicată (pentru ştiinţe tehnice şi economice)***, Partea I-a şi Partea II-a, Editura „Dimitrie Cantemir", Târgu Mureş, 2000, 414 p., ISBN 973-98920-6-x.
17. ***Managementul cercetării***, Editura „Dimitrie Cantemir", Târgu Mureş, 2000, 288 p. ISBN 973-8042-26-7.

Material cu drept de autor

18. *Tehnologia fabricației produselor industriale*, Târgu Mureş, Editura "Dimitrie Cantemir", 2001, 233 p., ISBN 973-8042-27-5.
19. *Economia mediului*, Târgu Mureş, Editura "Dimitrie Cantemir", 2001, 290 p., ISBN 973- 99596 – 6 – 0.
20. *Analiza valorii*, Târgu Mureş, Editura "Dimitrie Cantemir", 2001, 363 p., ISBN 973 – 8042 – 09 – 7.
21. *Analiza valorii*, Iaşi, Casa de Editură Venus, 2001, 295 p., ISBN 973 – 8174 – 38 – 4.
22. *Ecologie spirituală*, Iaşi, Casa de Editură Venus, 2002, 297 p., ISBN 973 – 8174 – 46 – 5.
23. *Statistica (probleme)*, Iaşi, Casa de Editură Venus, 2002, 144 p., ISBN 973 – 8174 – 49 – X.
24. *Statistica*, Iaşi, Casa de Editură Venus, 2002, 307 p., ISBN 973 – 8174 – 66 – X.
25. *Statistica*, Iaşi, Casa de Editură Venus, 2003, 371 p., ISBN 973 – 8174 – 85 – 6.
26. *Ecologie spirituală*, Iaşi, Editura PIM, 2003, 306 p., ISBN 973 – 7967 – 36 – 4.
27. *Statistica*, Iaşi, Editura PIM, 2003, 384 p., ISBN 973 – 7967 – 39 – 9.
28. *Statistica (probleme)*, Iaşi, Editura PIM, 2003, 210 p., ISBN 973 – 7967 – 50 – 2.
29. *Economia mediului*, Iaşi, Editura PIM, 2004, 316 p., ISBN 973 – 7967 – 74 – 7.
30. *Ecologie spirituală*, Iaşi, Editura PIM, 2004, 312 p., ISBN 973 – 716 – 036 – 3.
31. *Econometrie*, Iaşi, Editura PIM, 2007, 294 p., ISBN 978-973 – 716 – 603 – 6.

Pentru calitatea activității didactice desfăşurată, prin ordinul ministrului nr. 7626 din 15 iunie 1987, mi s-a conferit titlul de "CONFERENȚIAR UNIVERSITAR EVIDENȚIAT".

În 1987 am primit Premiul "Aurel Vlaicu", acordat de Academia Română pentru lucrarea "***Rulmenți. Proiectare şi tehnologie***", cu Diploma nr. 67 din 4 decembrie 1987.

De-a lungul timpului am avut diferite activități cu caracter obştesc de interes general pentru colectivitățile umane din care am făcut parte.

Între 1961-1964 am fost preşedintele Consiliului Uniunii Asociațiilor Studenților din Institutul Politehnic Iaşi.

Între 1969 şi 1976 am făcut parte din Consiliul tehnico-economic al Întreprinderii de Rulmenți Bârlad. Între 1977-1979 am făcut parte din Consiliul Oamenilor Muncii al Întreprinderii de Utilaje şi Piese de Schimb Botoşani, iar între 1979-1987 am făcut parte din Consiliul Oamenilor Muncii de la Întreprinderea Metalurgică Iaşi. În aceste calități am făcut parte din comisiile de prognoză şi cercetare ştiințifică a unităților economice respective, contribuind la stabilirea priorităților privind asimilarea progresului tehnic pentru produsele realizate sau/şi procesele tehnologice utilizate în unitățile economice menționate.

Material cu drept de autor

Am participat, temporar, în diferite comisii tehnico-economice având ca scop dezvoltarea tehnico-economică la diverse unități economice cum ar fi: Întreprinderea de Utilaj Greu (CUG) Iaşi, Întreprinderea Mecanică "Nicolina" Iaşi, Întreprinderea de Material Rulant Paşcani şi altele.

Aşa cum am mai menționat, între 15 septembrie 1978 - 1 octombrie 1990 am fost şeful Catedrei de Tehnologia Metalelor de la Institutul Politehnic Iaşi având în răspundere organizatorică şi îndrumare ştiințifico-didactică profilul metalurgic înființat atunci la Facultatea de Mecanică, cu patru specializări: ***Tehnologia turnării; Tehnologia deformării plastice la cald şi tratamente termice; Utilaj tehnologic pentru turnarea metalelor; Utilaj tehnologic pentru deformare plastică şi tratament termic*** (învățământ de zi şi seral), cu circa 1500 de studenți.

Între 1987-2004 am fost membru în Comisia Ştiința Materialelor a Academiei Române şi Preşedinte al Subcomisiei Ştiința Materialelor de la Academia Română - Filiala Iaşi.

Am fost organizator al diferitelor sesiuni ştiințifice pentru cadre didactice şi cercetători din unități de cercetare şi producție. Am făcut parte din diferite jurii naționale ale Conferințelor sau Simpozioanelor Naționale ale Cercurilor Ştiințifice Studențeşti şi membru ale unor Comisii ale M.E.I. de analiză a învățământului universitar în profil mecanic şi metalurgic. Am făcut parte, mai mulți ani, din comisia de admitere a Institutului Politehnic "Gheorghe Asachi" Iaşi şi de la Facultatea de Mecanică. Am fost membru în comisii de elaborare de subiecte pentru examenul de admitere în facultate.

Am făcut parte din diferite comisii ale M.E.I. pentru elaborarea sau îmbunătățirea de planuri de învățământ, programe analitice cadru, programe de perspectivă pentru dezvoltarea învățământului în România.

Din 1987 sunt expert tehnic pe lângă Tribunalul Iaşi.

Sunt coautor la următoarele invenții:

Certificat de Inventator nr.86.463 din 15.01.1985 pentru: "***Dispozitiv de superfinisare***".

Certificat de Inventator nr.92.850 din 27.05.1987 pentru: ***"Aparat pentru determinarea gradului de texturare a tabelelor***".

Certificat de Inventator nr.95.467 din 18.03.1988 pentru: ***"Procedeu de obținere a fontelor cu proprietăți fizico-mecanice superioare"***.

Certificat de Inventator nr.96.3312.11.1986 pentru: ***"Cap de forjare orbitală"***.

Posed Atestat editorial nr. 543 din 18.VI.1992, eliberat de Ministerul Culturii.

Ca urmare afirmării pe linie ştiințifică sunt menționat în:

Dictionary of Interantonal Biography, volume XVIII, publication October 1983, Cambridge, England.

International Who's in Who in Engineering, 1982/1983, Cambridge, England.

5.000 Personalities of the World, Edition Two, 1987, Published by the American Biographical Institute.

Material cu drept de autor

The International Directory of Distinguished Leadership, American Biographical Institute, Inc., Millenium Edition, 2000.

Dicționarul specialiştilor. Un "WHO'S WHO" în ştiința şi tehnica românească. Vol.1, Bucureşti, Editura Tehnică, 1995.

Septembrie 2007

Material cu drept de autor

CUPRINS

	Pag
INTRODUCERE	3
Cap.1. VARIABILE STATISTICE	9
1.1. Concepte de bază folosite în statistică	9
1.2. Variabile dependente şi variabile independente	10
Cap.2. PROBABILITATE ŞI DISTRIBUŢII DE PROBABILITATE	12
2.1. Noţiunea de probabilitate	12
2.2. Funcţie de repartiţie	16
2.3. Densitatea de probabilitate.	20
CAP.3. LEGI CLASICE DE PROBABILITATE	23
3.1. Legea binomială de distribuţie	23
3.2. Legea hipergeometrică de distribuţie	28
3.3. Legea de distribuţie a lui Poisson	31
3.4. Legea normală de distribuţie	33
3.5. Legea normală normată a lui Laplace	39
3.6. Verificarea corespondenţei dintre repartiţiile teoretice şi cele empirice	44
Cap.4. POPULAŢIE STATISTICĂ ŞI EŞANTION STATISTIC	57
4.1. Cercetări selective: de la populaţie la eşantion	57
4.2. Reprezentativitatea eşantionului	57
4.3. Erorile cercetării statistice prin sondaj	58
4.3.1 Erori efective. Verificarea reprezentativităţii eşantionului	59
4.3.2 Eroarea medie probabilă şi eroarea limită	60
4.4. Determinarea mărimii eşantionului	66
4.5. Probleme privind prognoza volumului de eşantionare. Dispersii marginale	70
4.6. Determinarea volumului eşantionului pentru selecţia aleatoare simplă	73
Cap.5. PRELUCRAREA DATELOR STATISTICE	77
5.1.Analiza preliminară a datelor statistice	77
5.2. Criterii pentru eliminarea valorilor ce diferă semnificativ de restul selecţiei	79
5.3. Prelucrarea primară a datelor statistice	83
5.4. Serii de distribuţie a frecvenţelor	88
5.5. Indicatori ai tendinţei de grupare a datelor seriilor statistice	92

Material cu drept de autor

5.5.1. Indicatori de medii: media aritmetică, armonică, pătratică, geometrică 92
5.5.2. Indicatori de poziție 109
5.5.3. Momente 111
5.6. Indicatori de analiză ai tendinței de împrăştiere a datelor statistice 113
5.7. Inegalitatea lui Cebâşev 121
5.8. Estimarea parametrilor statistici 122
5.8.1. Considerații introductive 122
5.8.2. Estimarea prin interval de încredere 123
5.8.3. Estimarea unei proporții 123
5.8.4. Estimarea unei medii 125
5.8.5. Estimarea unei varianțe 126
5.9. Testarea ipotezelor statistice 127
5.9.1. Teste unilaterale şi bilaterale 127
5.9.2. Regiune de acceptare şi regiune critică 128
5.9.3. Alegerea unui test 128
5.9.4. Influența eşantionării 129
5.9.5. Testul mediei unei legi normale de abatere tip cunoscută 129
5.9.6. Testul mediei unei legi normale de abatere tip necunoscută 131
5.9.7. Test a unei varianțe de lege normală, media fiind cunoscută 132
5.9.8. Testul unei varianțe de lege normală, media fiind necunoscută 133
5.9.9. Testul unei proporții 133
5.9.10. Test între ipoteze compuse 134
A. Testul unei medii de lege normală, abaterea tip fiind cunoscută 134
B. Testul unei medii a legii normale, cu abaterea tip necunoscută 136
C. Testul unei varianțe a legii normale, media fiind cunoscută 136
D. Testul unei varianțe al legii normale, media fiind necunoscută 136
E. Testul unei proporții 136
5.10. Teste de comparație 137
5.10.1. Comparația a două medii 137

Material cu drept de autor

5.10.2. Comparația a două variante 139
5.10.3. Comparația a două proporții 139
5.11. Test de adecvare 140
5.12. Testul χ^2 140
5.13. Serii de distribuție bidimensionale 142
Cap.6. ANALIZA DISPERSIONALĂ 153
6.1. Considerații introductive 153
6.2. Criteriul de egalitate a două dispersii 155
6.3. Criterii de egalitate a unui şir de dispersii 156
6.4. Criterii ale egalității mediilor de selecție 157
6.5. Analiza dispersională (ANOVA) unifectorială 160
6.5.1. Analiza variației cu un criteriu de clasificare şi eşantionare randomizată 162
6.5.2. Comparație între testul F de analiza variației cu două grupe de date statistice şi testul t Student pentru două probe independente 172
6.5.3. Teste pentru omogenitatea variației între mai multe eşantioane: testele Hartley, Cochran, Bartlett 174
6.6. Analiza variației în populațiile divizate în grupe 180
Cap.7. CORELAȚIE ŞI REGRESIE 189
7.1. Considerații preliminare 189
7.2. Forme de legături existente între fenomene şi procese economico-sociale 189
7.3. Covarianța şi corelația 191
7.4. Coeficientul de corelație 191
7.5. Corelație simplă (liniară) 192
7.6. Regresia liniară simplă cu o singură variabilă independentă 197
7.7. Corelație şi regresie liniară multiplă 205
7.8. Raportul de corelație şi coeficientul de determinație 211
7.9. Regresia neliniară 217
7.10. Corelația neparametrică 221
CAP.8. SERII CRONOLOGICE (SCR) 239
8.1. Conceptul de serii cronologice 239
8.2. Clasificare seriilor cronologice (SCR) 239
8.3. Analiza seriilor cronologice de intervale 241
8.4. Analiza seriilor cronologice de momente 250
8. 5. Ajustarea seriilor cronologice 252

Material cu drept de autor

8.6. Analiza statistică a variațiilor sezoniere 264
CAP.9. METODA INDICILOR 271
9.1. Conceptul de indici statistici 271
9.2. Indici în formă de bază şi în formă de lanț 272
9.3. Clasificarea indicilor după funcțiile lor cognitive 274
9.4. Indicii agregați. Sisteme de ponderare folosite la construirea indicilor de grup 276
9.5. Principii de bază ale aplicabilității indicilor agregați 279
9.6. Indicii nivelurilor medii 288
9.7. Descompunerea pe factori a variației unui fenomen complex folosind metoda indicilor 291
9.8. Serii cronologice de indici statistici 301
9.9. Teste de verificare a indicilor 303
9.10. Indici teritoriali 304
9.11. Metode de ierarhizare a unităților spațiale 305
CAP.10. ANALIZA STATISTICĂ A UNOR FENOMENE ECONOMICE SPECIFICE 315
10.1. Analiza statistică a productivității muncii 315
10.2. Analiza statistică a dinamicii fondului de salarii 337
10.3. Analiza statistică a dinamicii salariului mediu 342
BIBLIOGRAFIE 355
ANEXE 358

Material cu drept de autor

Material cu drept de autor

Tipar Digital realizat la **Tipografia** pim
Șoseaua Ștefan cel Mare nr. 11
Iași - 700498
Tel. / fax: **0232-212740**
e-mail:editurapim@pimcopy.ro
www.pimcopy.ro

Material cu drept de autor